STRUCTURAL STUDIES OF PHAGE LYSIS PROTEINS AND THEIR TARGETS

A Dissertation

by

VLADIMIR BORISOVICH KUZNETSOV

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2011

Major Subject: Chemistry

Structural Studies of Phage Lysis Proteins and Their Targets

STRUCTURAL STUDIES OF PHAGE LYSIS PROTEINS AND THEIR TARGETS

A Dissertation

by

VLADIMIR BORISOVICH KUZNETSOV

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	James Sacchettini
Committee Members,	Ryland Young
	David Russell
	Paul Cremer
Head of Department,	David Russell

August 2011

Major Subject: Chemistry

ABSTRACT

Structural Studies of Phage Lysis Proteins and Their Targets.

(August 2011)

Vladimir Borisovich Kuznetsov, B.S., University of California, Santa Cruz

Chair of Advisory Committee: Dr. James Sacchettini

Bacteriophages (phages) are viruses that infect bacteria. The phages that are described by this dissertation encompass 2 classes, double-stranded DNA phages and single-stranded RNA phages. While both of these phages infect similar bacteria, they have adopted different mechanisms to lyse, or destroy, the cell in order to release phage progeny. dsDNA phages have large genomes (>20 kb) and use multiple lysis proteins (holin, endolysin, and spanin complex) to lyse the cell. ssRNA phages, on the other hand, have small genomes (<6 kb) and only encode one lysis protein.

The two X-ray crystallography projects outlined here deal with the phage proteins involved in these lysis mechanisms.

The project described in the first study deals with the holin (T) and the antiholin (RI) of the ds-DNA phage T4, the major players of the lysis inhibition (LIN) phenomenon. Crystal structures of the holin and of the holin-antiholin complex are presented. The structures provide new molecular level insights into the phenomenon of LIN in bacteriophage T4 and the T-even phages in general.

The second investigation describes ongoing efforts at structural characterization of A_2, the maturation protein of the ssRNA bacteriophage Qβ that inhibits *E. coli* MurA. In addition, the structure of *Bacillus subtilis* MurA, which is not recognized by A_2, is presented. The crystal structure of *B. subtilis* MurA, the first structure of MurA from a Gram-positive organism, allows for a direct comparison of Gram-positive and Gram-negative homologs and for identification of any significant structural differences. The more flexible catalytic loop of *B. subtilis* MurA protrudes farther out compared to the loop of *E. coli* MurA and creates enough hindrance to prevent A_2 from establishing secure contact points.

DEDICATION

To my mother

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. James Sacchettini for his continued support and guidance. It is greatly appreciated. I would also like to thank Dr. Ryland Young for the collaborations and for making me feel welcome during my stay in his lab. Finally, I thank Sacchettini and Young lab members, past and present.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION

X-RAY CRYSTALLOGRAPHY

Overview

In December 2010, the Protein Data Bank (PDB) reached a new milestone- the number of macromolecular structures available in the PDB archive exceeded 70,000. The PDB is the single, global archive for information about the 3D structure of macromolecules and their complexes, as determined by X-ray crystallography, NMR spectroscopy, and cryo-electron microscopy. The number of entries available was only seven when the PDB was founded, with great foresight, at the Brookhaven National Laboratory in 1971. More than 87% of all structures deposited in the PDB are determined by X-ray crystallography.

X-ray crystallography is one of the most exciting and influential areas of modern biology. The ability of biomolecules to form crystals that are able to diffract the X-rays is the foundation of protein crystallography. Bragg's law, describing the relationship between angles and phases of the incident and reflective waves along with the interatomic distances in a crystal matrix, allows the reconstruction of a 3D crystal

---

This dissertation follows the style of *Journal of Biological Chemistry*.

structure based on its diffraction data. Modern crystallography has come a long way since Max Perutz determined the molecular structure of hemoglobin in 1959. Hundreds of laboratories and pharmaceutical companies all over the world use X-ray crystallography in their research utilizing the latest technologies from the areas of physics, robotics, and computer programming. Thousands of newly solved structures are being deposited in the online database RCSB PDB annually. But despite the unquestionable progress of the last thirty years, X-ray crystallography remains a very labor-intensive process where simple luck still plays a role. There are presently no universal rules or guaranties that would yield a well diffracting crystal. Based on numerous factors, solving a crystal structure could take anywhere from a few months to years of research. The overall process of obtaining a three dimensional model of a biomolecule involves a number of steps requiring a certain proficiency in the fields of biochemistry, molecular biology, statistical data analysis, and software interfaces. These steps include cloning and overexpression of the genes encoding the target proteins, crystallization of the purified proteins, data collection and processing, and computer modeling.

Cloning and overexpression of proteins

A successful crystallization experiment usually requires a substantial amount of highly purified material for the initial screening, typically in the range of 10-100 mg. The vast majority of protein targets are present in the cell in such miniscule amounts that their direct extraction in sufficient amounts is impossible. It is hard to imagine the field

of X-ray crystallography to be where it is today without the development of recombinant DNA technologies in the 1970s. Using recombinant DNA techniques, microbial cells could be engineered to over-produce foreign proteins. A fragment of coding DNA is inserted in a bacterial vector under the control of a powerful promoter and the resulting plasmid is transformed into a suitable cell line for subsequent over-expression. One of the most common plasmids used for overexpression of proteins is the pET family of vectors from Novagen based on the T7 promoter-driven system. pET plasmids not only allow cloning of genes under the control of a very powerful promoter- T7 RNA polymerase, they also give an option to mark a protein with an affinity tag, for example a poly-histidine (His_6) tag which aids in protein purification. Not all overexpressed proteins could be obtained in soluble form. There are different ways to improve the solubility of a recombinant protein. Most popular ones include fusing the expressed protein with another well-folded soluble protein such as maltose-binding protein (MBP) or truncating the protein to relevant sub-domains. Often times slowing the rate of protein production by lowering inducer concentration and/or growing the cells at lower temperatures (below 25° C) helps correct protein folding resulting in a more soluble protein.

Protein purification

In general, the purer the protein, the higher are the chances of obtaining a well-diffracting crystal. Usually, at least a couple of chromatographic steps are necessary to achieve the purity of 95% or higher. The first step of any protein purification is the

lysing of cells either via ultrasound (sonication), or by applying excessive pressure (French pressure cell press). Clarified cell lysate is then applied to an affinity chromatography column if the target protein is outfitted with an affinity tag, such as His_6-tag. It is common to use ion exchange chromatography, which separates biomolecules based on their net charge, as an intermediate step in the purification. Size exclusion chromatography, which separates proteins by size, is used as the final polishing step - to remove remaining impurities and to resolve oligomeric stages of the protein. Once the protein is deemed sufficiently pure and homogenous, it is concentrated to 10-50 mg/ml; its aliquots may be flash frozen in liquid nitrogen, and stored at -80° C.

Protein crystallization

Proteins are challenging molecules to crystallize – they are large, flexible and often presented as a non-uniform population in solution. Protein solubility should be gradually decreased in a controlled way to induce crystal formation. In a crystallization experiment, relatively concentrated protein solution is mixed with reagents called precipitants. As precipitating reagent composition is individual for each protein, a large number of conditions have to be probed (screened) to find one producing a crystal. A typical set of screened conditions includes polymers, salts, and buffers. A small drop of protein mixed with precipitation solution is left to gradually concentrate by means of various methods: vapor diffusion, dialysis, or micro-batch. The hanging and sitting drop vapor diffusion methods are the most popular. To maximize the likelihood of successful crystallization, the protein is screened against hundreds of crystallization conditions

covering a wide range of pH, chemical species, and concentrations. Well-equipped laboratories use robotic setups of nano-drops, such as Mosquito robot, to maximize the number of screenings. Once an initial crystallization condition is found, it often needs to be optimized in order to improve crystal quality. Optimization is done by screening around the conditions- varying protein to precipitant ratios, buffer pH, and temperature of crystallization, and by including various additives, for example, detergent in the solution.

Data collection, processing, and model building

Protein crystals are soft and fragile, with an average solvent content of 50%. They are thus very susceptible to the ionizing radiation of X-rays and have to be collected at cryogenic temperature. To prevent ice formation upon flash-freezing a cryoprotectant must be used most of the time. Crystallographic data are collected either at home using Rigaku or Bruker X-ray diffractometers with Cu rotating anode radiation source, or at synchrotrons, the world most powerful sources of coherent X-ray radiation. Collecting data at a synchrotron has a number of advantages: the X-ray beam is hundreds times more intense than the one on a home source, and the beam is tunable, a requirement for crystallographic experiments involving experimental phasing. In the data collection experiment only the intensities (amplitudes) of the diffraction are captured by a CCD (coupled charge device) or an image plate detector, while the phase relations between the reflections are lost. This phenomenon is known in crystallography as the phase problem.

There are two major avenues for reconstructing the missing phases. If a homology model for another protein with at least 40% of sequence identity already exists, molecular replacement technique is often enough to produce the initial phases for the reconstruction of the electron density. Otherwise, some additional phasing experiments must be performed with heavy atoms introduced into crystals. One of the most popular experimental phasing techniques relies on anomalous dispersion (SAD or MAD) of the so-called “heavy atoms”. One of the most common choices for introducing a heavy atom into protein crystal involves replacing methionine residues with seleno-methionine during protein over-expression. The Se atoms then act as heavy atoms covalently incorporated into the protein crystal.

The collected reflections are first indexed, i.e. crystal symmetry (or, the space group of the crystal), and the dimensions of the unit cell are determined. After the data are integrated, they are merged to account for any redundancy, and scaled for consistency in the intensity. The scaled data set is used to produce the initial phase estimates by aforementioned techniques. The inverse Fourier transform is then performed to reconstruct the electron density from the diffraction data.

Once the electron density map is obtained, a model of the structure is being built into it. The model then undergoes multiple cycles of additional rebuilding and refinement to achieve the best possible agreement between the observed and calculated structure factors. During refinement, the model is subjected to an impressive display of validation tools ranging from geometry checks to folding restrictions. The finished model is saved as a .pdb file that contains three-dimensional coordinates of all its atoms.

All new models are deposited to www.rcsb.org and are available to anyone with an Internet access.

BACTERIOPHAGES AND PHAGE LYSIS PROTEINS

Overview

Bacterial viruses, or bacteriophages, are the most widely distributed and diverse entities in the biosphere. Anywhere in the world where bacteria could survive, bacteriophages are present. British scientist Frederik Twort discovered bacteriophages in 1915. Few years later, in 1917, a French-Canadian microbiologist Felix d'Herelle made the same discovery, independently. Since then, bacteriophages have been widely studied throughout the world. Scientists working with phage made a number of critical discoveries in the area of biology such as the discovery that DNA is the carrier of genetic instructions.

Phage lysis proteins and their targets

The phages that are described by this dissertation encompass 2 classes, double-stranded DNA phages and single-stranded RNA phages. While both of these phages infect similar bacteria, they have adopted different mechanisms to lyse, or destroy, the cell in order to release phage progeny. dsDNA phages have large genomes (>20 kb) and use multiple lysis proteins (holin, endolysin, and spanin complex) to lyse the cell. ssRNA phages, on the other hand, have small genomes (<6 kb) and only encode one

lysis protein (1). The two X-ray crystallography projects outlined below deal with the phage proteins involved in these lysis mechanisms.

Holin and antiholin of phage T4

The first project deals with the holin (T) and the antiholin (RI) of the ds-DNA phage T4, the major players of the lysis inhibition (LIN) phenomenon. Lysis inhibition of T4-infected cells is a phenomenon that was identified in the 1940's and used for the discovery of fundamental genetic processes (2). Recently, it has been shown that LIN involves an interaction between the periplasmic domains of holin, T, and antiholin, RI of T4. Although the structure of a holin has been solved to intermediate resolution by cryo-EM (3), no atomic resolution structures of holin or antiholin was available. The information derived from the crystal structure determination of the antiholin RI, as well as the T—RI complex described in Chapter II adds an important tool in understanding the process of lysis inhibition (LIN) and of holin-dependent lysis in general.

A_2 of phage Qβ and *B. subtilis* MurA

The second project, described in Chapter III, outlines an ongoing effort at the structural characterization of A_2, the maturation protein of the ssRNA bacteriophage Qβ that inhibits *E. coli* MurA, the enzyme which catalyzes the first committed step in the cell wall biosynthesis pathway (4). An atomic resolution structure of the A_2—MurA complex will further our understanding of the underlying inhibition mechanism and pave the way for the design of novel antimicrobial peptides.

In addition, the structure of *Bacillus subtilis* MurA, which is not recognized by A_2, is analyzed. It has been recently shown that A_2 does not associate with *B. subtilis* MurA *in vitro* (5). Since the structures of *E. coli* MurA in both open and closed conformations are available, the structure of *B. subtilis* MurA, the first from a Gram-positive organism, allows for the direct comparison of Gram-positive and Gram-negative homologs and identifies any significant structural differences. It also helps in explaining the specificity of A_2 against *E. coli* MurA.

CHAPTER II

STRUCTURAL STUDIES OF T4 LYSIS PROTEINS

BACKGROUND

Viruses, from the Latin *virus* meaning poison, are ever-present parasites of bacteria, plants, and animals. Viruses are much simpler organisms than cells; technically not classified as living, viruses lack the metabolic machinery to reproduce outside their host cells. A virus consists of a genome, made of RNA or DNA, stored inside a protein or protein-lipid shell called capsid. The shell is necessary to protect the genome during the transmission of virus from one host to another. In addition, the shell uses its sensors to locate suitable hosts and is responsible for the delivery of the viral genome into the host. Bacterial viruses are known as bacteriophages or phages.

Enterobacteria phage T4 is a double-stranded DNA phage that infects *E. coli* bacteria. T4 is a myophage; a class of phages characterized by a long contractile tail- the most frequent bacteriophage group. While phage Qβ has one of the smallest genomes, T4 has one of the largest, 172 kbp, which comprises 102% of the unique region of 169 kbp (Figure 2.1) (6). Thus, every T4 virion carries dsDNA with a 3.4 kb terminal redundancy (7). A sample of T4 DNA molecules is circularly permuted which makes the genetic map of T4 circular despite the fact that its genomic DNA is linear. The dsDNA chromosome is encased in an icosahedral head that is 115 nm long. T4 is one of the most complex viruses; its genome has 274 open reading frames with more than 40 of those

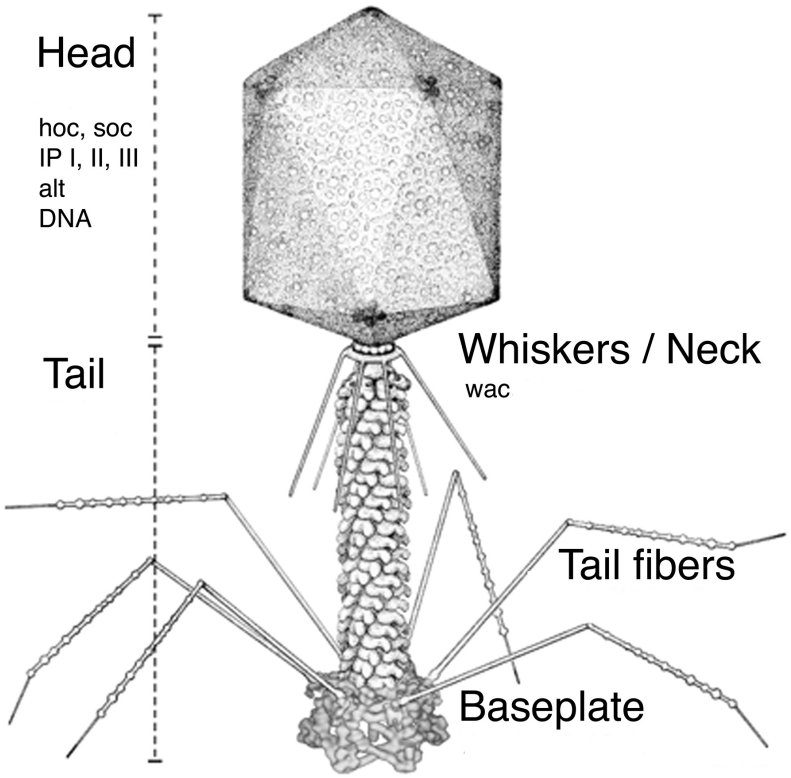


Figure 2.1 Structure of bacteriophage T4 (modified from (6)).

encoding structural proteins. A unique feature of T4 phage is its DNA composition. The DNA has had all of its cytosine residues replaced by hydroxymethylcytosine, which serves to inhibit Type I restriction systems of most strains of *E. coli*. Additionally, T4 is one of the largest phages at approximately 90 nm wide and 200 nm long. Phage T4 is made up of a prolate head with hemi-icosahedral ends (8) encapsidating the genomic DNA, a 100-nm long cocylindrical contractile tail that ends with a 46 nm-diameter baseplate (9), and six 145 nm-long fibers that are attached to the baseplate. Tailed bacteriophages in general, compared to animal viruses, are extremely efficient in infecting the host cells (7). Usually only one phage T4 particle is needed to infect a host cell. This high efficiency is attributed to the phage adsorption apparatus, which consists of the long tail fibers (LTFs), short tail fibers (STFs), the baseplate, and the contractile tail (6).

T4 bacteriophage initiates interaction with an *E. coli* cell when the phage tail fibers contact a binding site on a surface of the cell. Adsorption on the bacterial surface is crucial for infection. Infection starts when the T4 phage recognizes either the lipopolysaccharide cell surface receptors or OMPs of the host with its long tail fibers. This still reversible interaction supposedly allows the phage to move about the surface of the *E. coli* bacterium and produces the recognition signal that is eventually transmitted to the baseplate through the LTFs (6). The short tail fibers unfold from under the baseplate, where they are initially assembled, and bind to the lipopolysaccharide cell surface receptors, this time making the adsorption irreversible. Conformation of the baseplate changes from hexagonal to star, or open hex, (9), causing the tail sheath to contract

irreversibly to 36 nm long down from the original length of 100 nm. The tail tube on the other hand does not contract, protruding from the tail sheath and the baseplate, and punctures the outer membrane of the cell (10). When it reaches the periplasmic peptidoglycan layer of the host cell, the β-helix needle of the phage used to puncture the outer membrane dissociates from the tip of the tube activating three lysozyme domains (11). The lysozyme domains digest the peptidoglycan layer making penetration of the tail tube to the inner membrane of the cell possible. The phage DNA is ejected into the host via the tail tube (12).

Ejection of the phage DNA into the host initiates the lytic cycle. Phage enzymes completely shut down the cell's own DNA and RNA synthesizing pathways and replicate its own genome (7). Aforementioned uniquely decorated phage DNA allows T4 nucleases to distinguish between its own genomic DNA and that of the host. The host RNA polymerase is used exclusively for T4 transcription (13). Bacteriophage components begin to be produced by way of the host's metabolic machinery in the amount sufficient to assemble up to 200 new phage particles (37° C, 25 min.). The 20 or so phage are assembled using mononucleotides obtained from the degraded host chromosome. Morphogenetic assembly consists of three independent pathways: filled head, tail, and LTFs are made separately via ordered pathways (14). The phage DNA is packaged into a head in a spindle like fashion by the packaging apparatus powered by ATP hydrolysis (15,16). The mature phage particles are assembled inside the host, and at 25 minutes, the infection ends with the lysis of the host cell. The newly made phage are released into the environment.

Bacteriophage T4, like all double-stranded DNA phages, effects host lysis with a holin (T), an endolysin (E), and a spanin complex (pseT.3 and pseT.2) (17). Endolysin is a term for a class of enzymes whose purpose is to degrade the cell wall. The T4 endolysin, E is a "true lysozyme"; just like the egg white lysozyme, it hydrolyzes the β(1,4)- glycosidic bonds in the murein. The crystal structure of E is known (18). The holin is a membrane protein. T and E accumulate in the cytoplasmic membrane and cytoplasm, respectively, throughout the period of late gene expression. E accumulates, fully folded, but does not act until released into the periplasmic space. At an allele-specific time, T triggers to disrupt the membrane by creating a large lesion, allowing E to enter the periplasm and degrade the murein layer of the cell envelope. It is the holin then that determines the precise timing of lysis. The spanin complex tethers the IM and OM and serves to disrupt the OM upon degradation of the murein layer (19).

T4, like all known T-even phages, exhibit a unique phenomenon when dealing with the host cells (20). In single infection, cells lyse after about 25 minutes (LB, 37° C) releasing 100-200 new phage per host cell (20). Newly released phage now infect surrounding *E. coli* cells already infected by phage. The process of continuously infecting the cell with new phage particles is termed superinfection. In the event superinfection happens 3-5 minutes after the original infection, the host becomes lysis inhibited. The phenomenon is called lysis inhibition (LIN). The host is not lysed for up to several hours. When lysis does occur, the number of new phage accumulated inside a single cell could reach a thousand (20). The LIN state is unstable; it requires continuous

superinfection to be maintained. LIN state can be reversed by addition of energy poisons that cause the collapse of the membrane potential.

The phenomenon of lysis inhibition results in wild type T4 making small and fuzzy edged plaques on agar lawns because the cells on the perimeter of the plaque are in continuous state of LIN. This makes it possible to recognize T4 mutants defective in LIN since their plaque morphology is different- large and sharp edged (21). Mutants defective in LIN became known as T4*r*, where "*r*" stands for rapid lysis. Such mutants were mapped to multiple T4 loci. The *r* system was instrumental in elucidating some of the most fundamental principles of molecular genetics (22). Eventually, only two genes- *rI* (23) and *rV* (24) were found to be required for establishing LIN in all *E. coli* strains K-12, B, and K-12(λ) and for preserving the wild type plaque morphology. *rV* was later shown to be allelic to *t* (20), which encodes the holin T. Recently, Ramanculov et al. demonstrated that at the molecular level, the LIN phenomenon is due to the inhibition of the holin, T, by the antiholin, RI (25).

Bacteriophage holins are a diverse group of proteins. T4 holin has an unusual topology as compared to holins from other dsDNA phages (Figure 2.2). All other characterized holins have either three (e.g. S105 of phage λ, the most widely studied holin to date) transmembrane domains (TMDs) and belong to class I, or two (e.g. $S^{21}68$ of phage 21) transmembrane domains and belong to class II (17). There have been more than fifty unrelated families of holins from class I and class II identified as of 2000. In contrast, T has only one TMD and belongs to a unique class III with the N-terminus inside and the C-terminus outside of the inner membrane. Holin T, at 218 residues, is

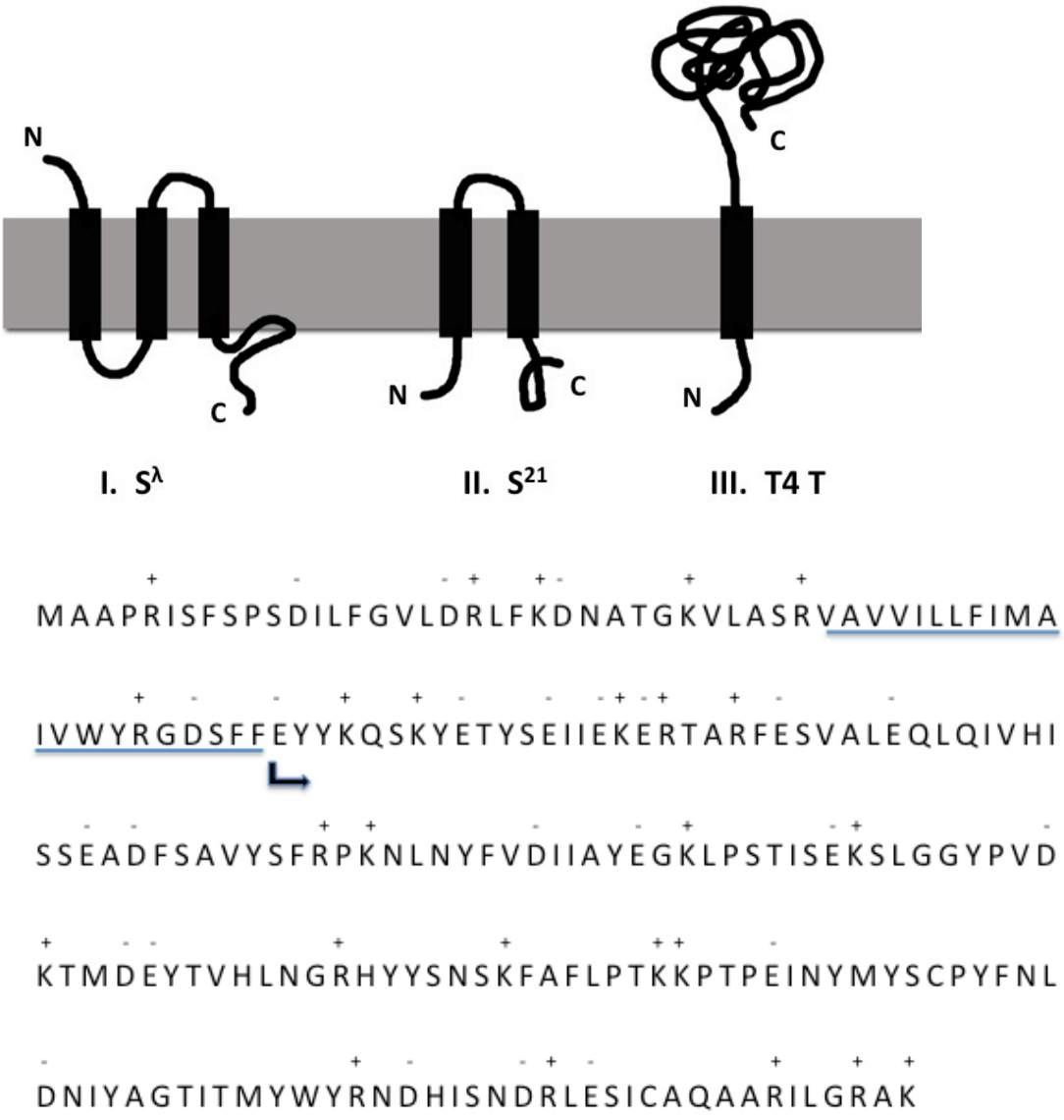


Figure 2.2 Topology of Class I, II, and III holins. Sequence of phage T4 holin T with the single transmembrane domain underlined.

also substantially larger than other holins (e.g. the λ holin S105 is only 105 residues). The extra mass comes primarily from its large C-terminal periplasmic domain of 163 residues, whereas other holins have no significant periplasmic component.

Holes or lesions created by T were visualized by C. Savva and S. Moussa (unpublished) using cryo EM methods. To examine the membrane lesions caused by T4 T, cells expressing the T holin were plunge frozen in liquid ethane and examined by cryo-electron microscopy. Approximately 1/3 of the cells observed showed lesions in the inner membrane with an average hole diameter of 200 nm. Most cells only exhibited one hole. The size and distribution of holes were similar to the ones formed by the lambda holin S105 (26). For S105, holin expression alone at near physiological levels leads to formation of large membrane lesions that measure an average of 340nm. 1-2 holes per cell were observed, but based on the geometrical constraints of viewing a hole of that size in the microscope, an average of 2 holes per cell were estimated. Some holes were as big as 1 micrometer. The holes were randomly distributed in the cell periphery. Tomography of some of the lesions indicated that they were of irregular shape and confirmed that the membrane gaps observed in projection were indeed holes rather than membrane deformation or invagination. Two possible models were discussed. One model, called “the Swiss cheese” model discusses smaller rings making small holes of 23nm that coalesce into bigger ones (3).

The second model was proposed by R. White et al. (27). She showed that GFP-fused holin accumulates in the cell membrane throughout the periphery up until the point of triggering and within a minute it forms discrete patches of fluorescence, termed the

rafts. She also showed that the protein in the rafts does not move after triggering but is mobile before triggering. The A52V mutant of holin S, which is blocked at the dimer stage, did not form rafts, whereas the A52G mutant that causes early lysis formed small rafts at an earlier time.

At this point it is impossible to predict whether the Swiss cheese or the rafts model will ultimately prevail. Proponents of the raft model argue that it explains how mutations in all 3 TMDs of S affect the holin so there must be some interaction on all sides of the helices. Also a raft model excludes lipid rather than extrudes it. But some are not convinced that the fluorescence data prove the existence of rafts citing inadequate resolution of the method. In cryo, the end result could be seen as one that may have started off as a raft or as small holes. Correlative cryo-fluorescence and cryo-electron microscopy should be able to tell if the rafts correspond to the large holes that were observed. Some early experiments by C. Savva (personal communication) suggest that they do not correlate, but that could mean the GFP-holin is not a good hole former and that is why the holes did not correlate with the fluorescence presence.

Recently, Tran et al. showed that lysis inhibition requires interaction of the periplasmic domains of T and RI (17).

The T4 antiholin, RI, is only 97 residues and contains an N-terminal hydrophobic domain and the 75-residue periplasmic domain. RI has two essential cysteines in the periplasmic domain that form an intra-disulfide molecular bond. The cysteines are conserved for all known antiholins of T-even phages (Figure 2.3). Gene *rI* is served by both early and late promoters (28) while *t* is only served by late promoters.

The N-terminal sequence of RI comprises a signal anchor release (SAR) domain, which anchors RI in the inner membrane but allows its release into the periplasm, without proteolytic processing (25). It is thought that the relatively high content of weakly hydrophobic and uncharged polar residues in the SAR domain allows it to exit the membrane. Unless the LIN state is established, RI is inactivated and subsequently degraded by the periplasmic protease, DegP. The SAR domain of RI has been identified as the major determinant of its instability and the target of DegP. Tran et al. showed that DegP is required for the rapid turnover of RI (17). The still unidentified signal for LIN probably comes from a superinfecting phage since T4 "ghosts", phages with heads emptied of DNA and internal proteins, fail to impose LIN (7). It is thought that small cytoplasmic membrane protein Imm (for immunity) produced as early as 1-2 minutes after the primary infection prevents any genetic material from the superinfecting phages from reaching the cytoplasm. In addition, the spackle protein was shown to play a role in changing the cell envelope to offer resistance to superinfecting phage. Both the DNA and more than 1000 internal proteins from a superinfecting phage are localized in the periplasm and could potentially serve as a signal for lysis inhibition received by RI. One of the working theories that the signal is DNA from a superinfecting phage is being currently tested by gel shift assays and NMR relaxation techniques by S. Moussa (Young lab, unpublished).

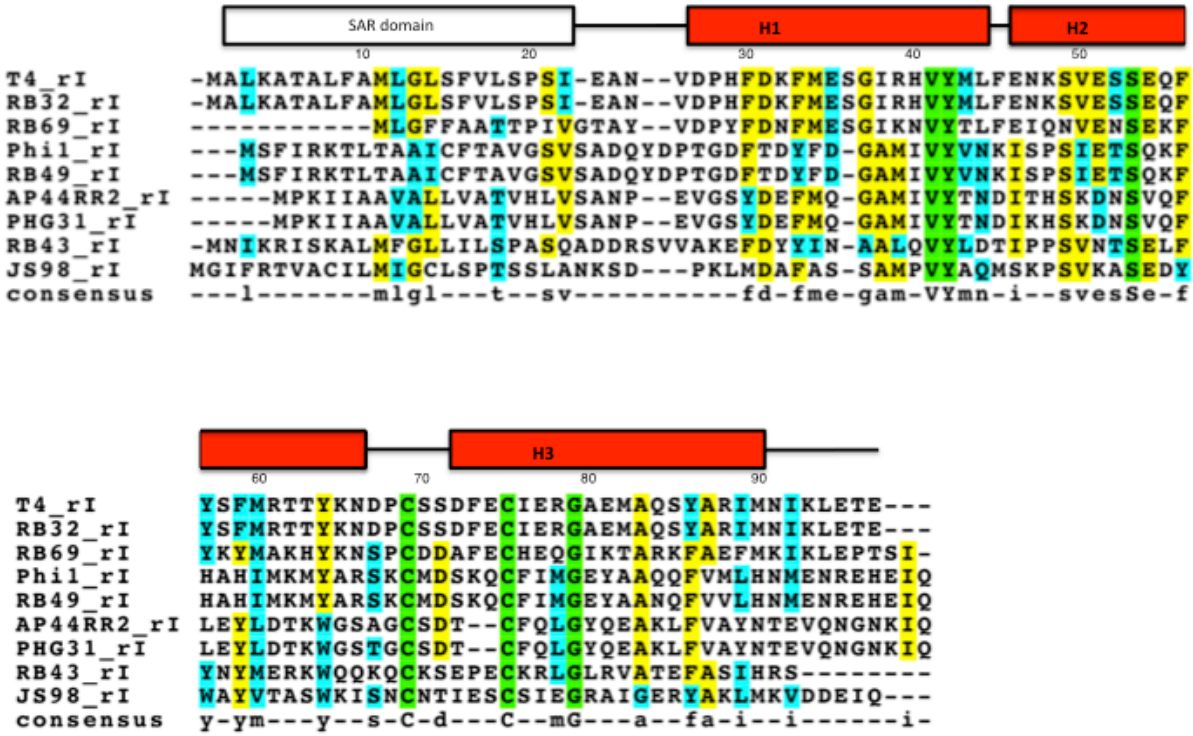


Figure 2.3 Sequence analysis of RI orthologs (T-even phages). Green indicates all match; yellow residues are conserved; blue residues are similar. Secondary structure elements are for RI of phage T4 (based on results below).

METHODS

Overexpression and purification of T4 RI soluble domain (sRI)

Full length RI is an unstable protein with the proteolytic half-life of 2-3 minutes (25). The SAR domain is a major determinant of RI instability. A truncated allele, *srI*, which encodes the soluble C-terminal domain of RI, was cloned into pET11a vector (NEB) (a T7 promoter driven protein expression system). The C-terminal domain of sRI has a non-cleavable poly-histidine tag. The periplasmic region of RI contains two essential cysteines; hence the use of a cell line that enhances disulfide bond formation in the cytoplasm is desirable. Origami2 (DE3) cell line from EMD was determined to be suitable by S. Moussa (Young lab). Origami 2 host strains are K-12 derivatives that have mutations in both thioredoxin reductase (*trxB*) and glutathione reductase (*gor*) genes, which greatly enhance disulfide bond formation in the *E. coli* cytoplasm. Recently, it was shown by S. Moussa that the yield of sRI could be doubled using the Shuffle (DE3) cells (NEB) designed with additional properties to enhance the disulfide bond formation and proper folding of the target protein.

The cells were grown in LB (Luria Broth) media by a modified procedure: after growing the culture at 37º C to a mid-log phase of $O.D._{550}$ 0.5-0.6, the flasks were transferred to an ice bath and swirled until the internal temperature reached 16º C. The flasks then were transferred back into the shaker pre-incubated at 16º C; culture was induced with 1mM IPTG, and grown for 18 hours. This low temperature induction resulted in >50% soluble protein (data not shown).

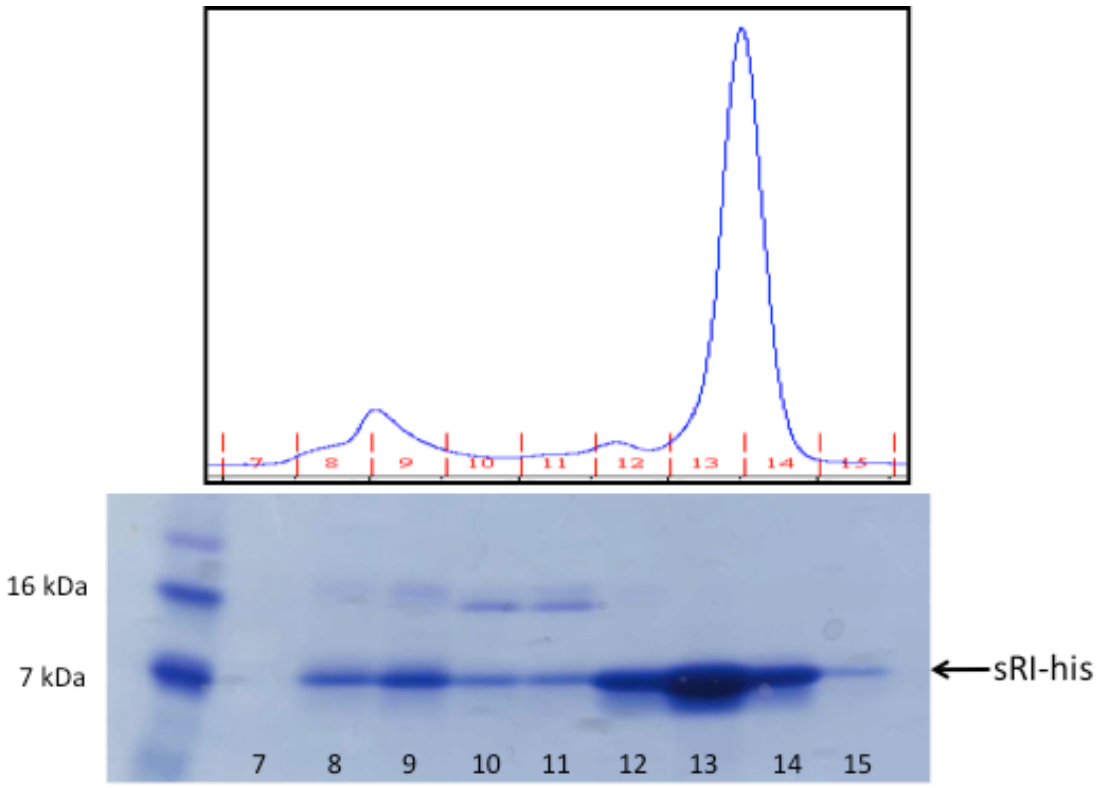


Figure 2.4 Purification of RI. Superdex 75 size exclusion chromatography of sRI. SDS-Page gel of the corresponding fractions is shown below (gel courtesy of S. Moussa).

The harvested cell pellet was resuspended in 20 mM phosphate buffer (pH 8.0) containing 100 mM NaCl and Protease Inhibitor Cocktail Set V (EMD). The cells were lysed by two passes through a cooled French pressure cell (Microfluidics M-110P) at 19000 psi. sRI was purified on an IMAC column (Talon Co^{2+} resin, Clontech) followed by the Superdex 75 (GE Healthcare) size exclusion chromatography (Figure 2.4). Protein was purified as a monodispersed monomeric fraction with purity >95% as judged by SDS-PAGE. The final yield was 3.5 mg/liter culture. The samples were concentrated immediately after purification to 10-13 mg/ml, flash frozen in liquid nitrogen without any cryoprotection, and stored at -80º C. To produce Se-Met labeled sRI, cells were cultured using methionine-free medium supplemented with 50 mg/liter L-Se-Met, and the protein was purified using the same protocol as for the native protein. Selenomethionine incorporation was confirmed by MALDI-TOF mass spectrometry.

It is important to acknowledge a significant input on the project from my collaborator Samir Moussa from Dr. Ry Young's laboratory. S. Moussa contributed all the molecular biology and genetic experiments, including creation, optimization and testing of the recombinant expression vectors, as well as the search for suitable bacterial cell lines. S. Moussa was also responsible for all mass spectrometry experiments mentioned in this chapter. In addition, he also helped prepare the recombinant proteins and in preparation of several of the crystallization trials, described below.

Crystallization of sRI

Initial crystallization screening was done at 16 °C on Phoenix robot (Art Robbins Instruments) using the sitting drop vapor diffusion method. Out of more than 700 crystallization assays screened, the two most promising results came from the in-house crystallization screen *Adventure Dive*. The *Adventure* series was designed by Inna Krieger (Sacchettini lab) to explore the crystallization space not already covered by commercial screens. The formulations are rich in additives, which aid in protein stabilization and potential crystal making contacts. The pH range of *Adventure* series is broader than the typical commercial matrices. Crystals were optimized in E3 condition from *Adventure Dive* (100 mM Na citrate (pH 4.5), 1M Na acetate, 0.5% Anapoe X114 detergent). Diffraction quality crystals were grown by combining equal ratios of the protein and reservoir solution by a hanging-drop, vapor diffusion method. The selenomethionine labeled derivative of sRI was crystallized under similar conditions.

Overexpression and purification of T4 T soluble domain (sT); purification of sT--sRI complex

sT, the periplasmic domain of T4 holin, T, comprises residues 56-218, with a non-cleavable his_6-tag at the C-terminal. sT-his was over expressed in *E. coli* under similar conditions to sRI-his. This resulted in a soluble protein that could be purified with a Talon (IMAC) resin. However, upon elution the protein oligomerized into a translucent jelly-like substance, probably reflecting the natural function of T to aggregate. The problem was rectified by S. Moussa who suggested eluting sT-his into

fractions containing a molar excess of purified sRI-his. The resulting complex was soluble and stable for at least a few days at 4º C. Size exclusion chromatography of the mixture on Superdex 75 column produced a significant peak corresponding to sRI—sT complex of ~50 kDa, separated from monomeric sRI-his and sT-his (Figure 2.5). The sample was concentrated to 8-10 mg/ml, flash frozen in liquid nitrogen, and stored at -80º C. Doubly (both sRI-his and sT-his) and singly (only sT-his) substituted selenomethionine derivatives of the complex were prepared.

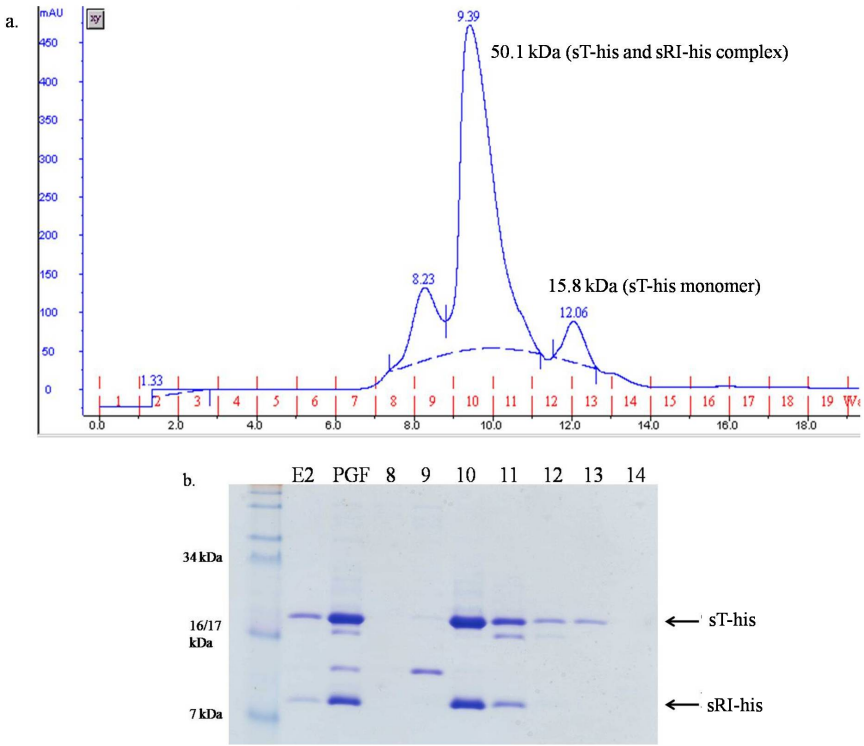


Figure 2.5 Purification of RI-T complex. Superdex 75 size exclusion chromatography of sRI-sT complex with corresponding fractions run on SDS-Page (gel courtesy of S. Moussa).

Crystallization of sRI-sT complex

Both the native and selenomethionine substituted sRI—sT complexes were crystallized. Initial crystallization screening was done at 16 °C by the sitting drop vapor diffusion method, using Mosquito (TTP Labtech) and Phoenix (Art Robbins Instruments) robots. Crystals of the native complex were grown by mixing 1.0 µl of the protein solution (10 mg ml^{-1} in 20 mM phosphate, pH 8.0, and 100 mM NaCl) and 1.0 µl of the reservoir solution (3.2 M NaCl and 0.1M Na acetate trihydrate, pH 4.6). Crystallization trials of the Se-Met labeled derivatives with above condition were unsuccessful. Crystals of the doubly Se-Met substituted complex were obtained by optimizing Index (Hampton Research) condition C8- 1.0 M ammonium sulfate, 0.1 M Bis-Tris, pH 5.5, and 1% PEG 3,350.

Data collection and structure determination

X-ray diffraction data were collected at 100 K on beamlines 19-ID and 23-ID at APS (Argonne National Labs, Chicago, IL). Additionally, Cr edge (λ=2.29 Å) data set of the Se-Met derivative of RI was collected at Rigaku America (Woodlands, TX).

Crystals of RI were cryoprotected in the reservoir solution supplemented with 24% (w/v) ethylene glycol. Diffraction data were processed using HKL2000 (HKL Research, Inc.). The structure of RI was determined by the SAD method using the 2.5-Å resolution data from the Se-Met labeled crystal. All of the potential six selenium atoms were located using SHELXD, and positions were further refined by Sharp. Initial automated model building was done within PHENIX suite of software programs. The

remaining 26% of the model was manually build using COOT and refined with PHENIX. Water molecules were manually added during iterative cycles of model building and refinement using an F_o-F_c map.

Due to extreme fragility of Se-Met crystals of RI-T complex, the cryoprotectant (30% ethylene glycol) was added directly to the sitting drop in 1:1 ratio. The experimental phases for the selenomethionine incorporated RI-T crystal were obtained using single-wavelength anomalous dispersion phasing. The crystal diffracted to a resolution of 2.3 Å. Diffraction data was processed using HKL3000. Automated structure solution was carried out by the PHENIX *AutoSol* wizard resulting in preliminary model. Model of RI-T was build by PHENIX *AutoBuild*. The model was further manually built using COOT and refined by PHENIX *Refine*. Water molecules were added in a similar fashion to RI model.

Data collection and refinement statistics are provided in Tables 2.1 and 2.2. Structural figures were prepared using UCSF Chimera. The structural similarity searches were made using VAST and DALI servers.

Table 2.1 Crystal Data and Refinement Statistics for sRI

Data Collection Statistics	
wavelength (Å)	0.97929
data set	peak
resolution (Å)	40.66-2.21
completeness % (last shell)	97.92 (95.1)
I/σ(I) (last shell)	
no. of unique reflections	
R_{sym} (%)	
Refinement Statistics	
unit cell dimensions	a=b=49.98 Å, c=118.68 Å
space group	P 6_2 2 2 (180)
no. of protomers per asymmetric unit (Z)	1
no. of reflections in the working set	4758
completeness (%)	97.9
R_{factor} (%) / R_{free} (%)	22.4 / 25.9
no. of protein residues	68
no. of solvent molecules	27
average B factor for protein (Å)	69
rmsd for bond length (Å)	0.008
rmsd for bond angles (deg)	0.988

Table 2.2 Crystal Data and Refinement Statistics for sRI-sT Complex

Data Collection Statistics	
wavelength (Å)	0.97947
data set	peak
resolution (Å)	44.39-2.50
completeness % (last shell)	99.64 (98.5)
I/σ(I) (last shell)	
no. of unique reflections	11577
R_{sym} (%)	
Refinement Statistics	
unit cell dimensions	a=b=120.13 Å, c=85.16 Å
space group	P 6_5 2 2
no. of protomers per asymmetric unit (Z)	1
no. of reflections in the working set	
completeness (%)	
R_{factor} (%) / R_{free} (%)	19.7 / 23.7
no. of protein residues	213
no. of solvent molecules	62
average B factor for protein (Å)	43.7
rmsd for bond length (Å)	0.009
rmsd for bond angles (deg)	1.04

RESULTS AND DISCUSSION

Overall structure of RI antiholin

The periplasmic domain of RI antiholin (sRI hereafter), composed of residues 25-97 with a noncleavable C-terminal $(His)_6$ tag, was over expressed and purified by Co-chelating affinity column followed by gel filtration. The single crystal of selenomethionine-labeled sRI diffracted up to 2.2 Å and the phases of the x-ray crystal structure were solved using the SAD method (Table 2.1). The crystal belongs to the hexagonal $P6_222$ space group with one protomer molecule in the asymmetric unit. The final model was refined with the R factor of 22.4 and R_{free} of 25.9 %, contains residues 26-92 of sRI and 27 water molecules. Residues 24, 25, and 93-97 are not observed in the electron density map and therefore omitted from the structure (Figure 2.6).

The single protomer of the asymmetric unit of sRI consists of three α-helices, exhibiting a long and kinked H1-H2 helix (Residues 28-44 and 46-66 respectively) and a C-terminal H3 helix (residues 72-90) (Figure 2.6). H2 is kinked with a bent-angle of 15 degrees at phe45 with reference to the N terminal helix H1. The helix H3 is connected to H2 through a hairpin loop, similar to helix-turn-helix structural motifs observed in the DNA binding proteins. The RI structure was searched against the Protein Data Bank using the DALI server and some notable homologous hits of longer lengths were found. The top scoring structures were helical regions of larger proteins of length ranging from 134-522 that only marginally resembled RI. Therefore, the structure of RI is unique, and there are no homologous structures that can give information about the function of RI.

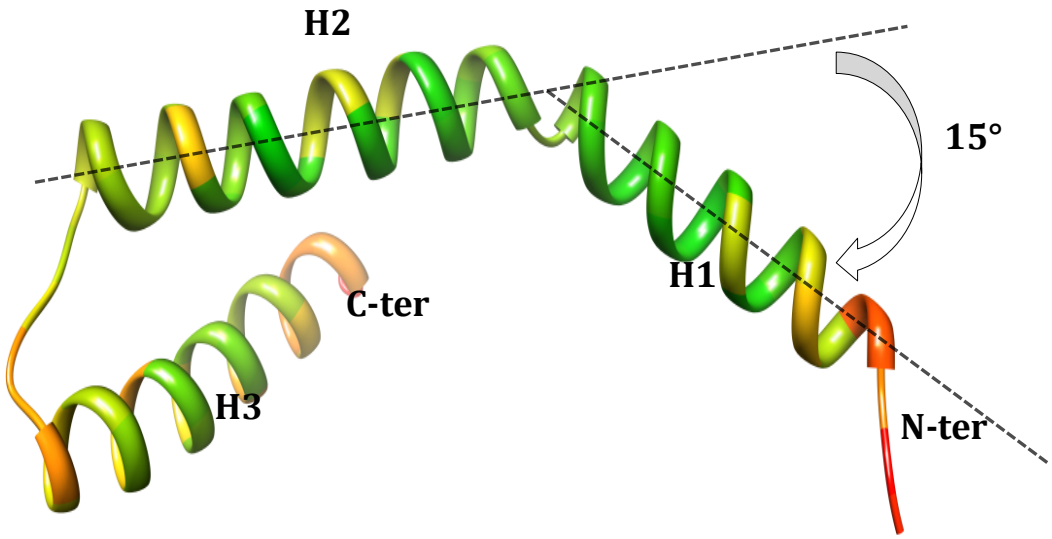


Figure 2.6 Overall structure of sRI. Residues are colored according to their B-factors. Long alpha helix is kinked at 15° according to HelixAng (ccp4). The angle between H2 and H3 is 130°.

Tetrameric assembly of RI is a dimer of dimers

Although the asymmetric unit of sRI contained only one protomer, a tetrameric assembly was clearly visible in the crystal lattice. The tetramer is composed of four identical symmetry related subunits as shown in Figure 2.7 (labeled A, B, C and D). The tetramer can be discussed as a dimer of dimers (AB dimer and CD dimer) based on the proximity of subunits as shown in Figures 2.7 and 2.8. The nonbonding interactions between subunits of AB dimer are visualized in Figure 2.9. In each dimer, sRI packs in a head to tail fashion in which helices H2 and H3 of one subunit (head) contacted the H1 helix (tail) of the other subunit; and vice versa. The dimer interfaces between chains A and B and chains C and D are strong enough to form a stable dimer: the buried surface area of the AB or CD dimer interface is 50 %. The free energy of dimer dissociation is 26.6 kcal/mol, suggested that the dimer is thermodynamically stable and the solvation free energy gain upon dimerization is -33.5 Kcal/mol. The amino acid residues of both the subunits contribute to several inter subunit interactions, within the dimeric assembly of RI. The helices H2 and H3 of the head region of each subunit are held together by intra subunit interactions. In addition, the conformation of the loop connecting the helices H2-H3 also favors this arrangement. The Cys69 of this loop forms a disulfide bond with the Cys75 of the H3 residue to lock itself in one conformation. The head region of one subunit, helices H2, H3 and the loop, together form a binding groove for the tail region of the next subunit. The dimer interface contains residues that form hydrogen bonds and hydrophobic interactions (Figure 2.8). The side chain atoms of Phe

30, Phe 33, Met 34, Ile 38, Val 41, Leu 44 and Phe 45, all from the N-terminal tail (H1 helix) of RI are involved in inter subunit hydrophobic interactions with the binding

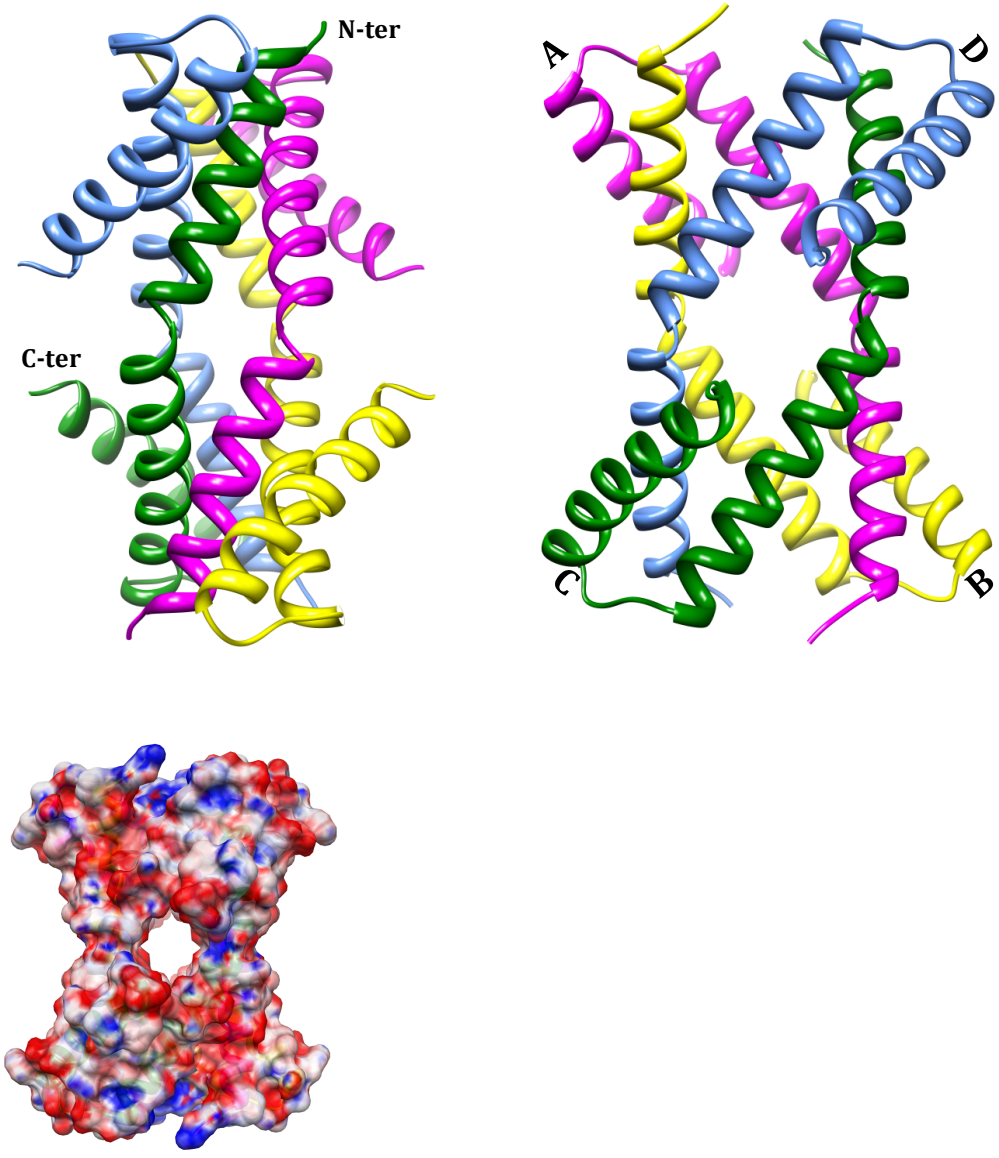


Figure 2.7 Tetrameric assembly of RI. Below is the electronegative surface of tetramer.

groove formed by the head portion of the next subunit. In addition, the OD1 atom of Asp27 (one subunit) makes a hydrogen bond with the hydroxyl group of the Tyr 64 (next subunit, 3.2 Å distance) and OB atom of Ser 36 hydrogen bonds with the main chain carbonyl oxygen atom of Ile 76 of the next subunit (at 2.6 Å distance).

Within the tetramer ($\Delta G(int)$ =-87.1 kcal/mol and $\Delta G(dis)$=13.8 kcal/mol) in the crystal lattice, the dimers hug each other through several dimer-dimer interactions as shown in Figure 2.8. In contrast to the head to tail interaction of the dimer, the dimer – dimer interface is mainly formed between the H1-H2 kinked helices of all four subunits. The long kinked helices H1-H2 of one subunit run side by side with the H1-H2 region of the next subunit (Figure 2.8). In each of these interfaces, the Tyr42 of one subunit points its side chain in to the kink of the next subunit and stacks with the side chain of Tyr42' of the next subunit. The side chain hydroxyl group of Tyr42 of one subunit also hydrogen bonds with the main chain carbonyl atom of Tyr42' of the next subunit.

Therefore, it is possible that sRI could exist as a dimer and/or a dimer-dimer. The gel filtration profile of RI showed one major peak near 8 KDa and a minor peak near 16 KDa suggesting RI could predominantly exist as a monomer in solution. However, a compact oligomeric assembly like RI might migrate in the smaller molecular weight range compared to other globular proteins. Further, the complementarity between inter subunit head-tail interactions suggest that the RI molecules should have a natural propensity towards a more stable dimer or tetramer form. It is possible that the simple

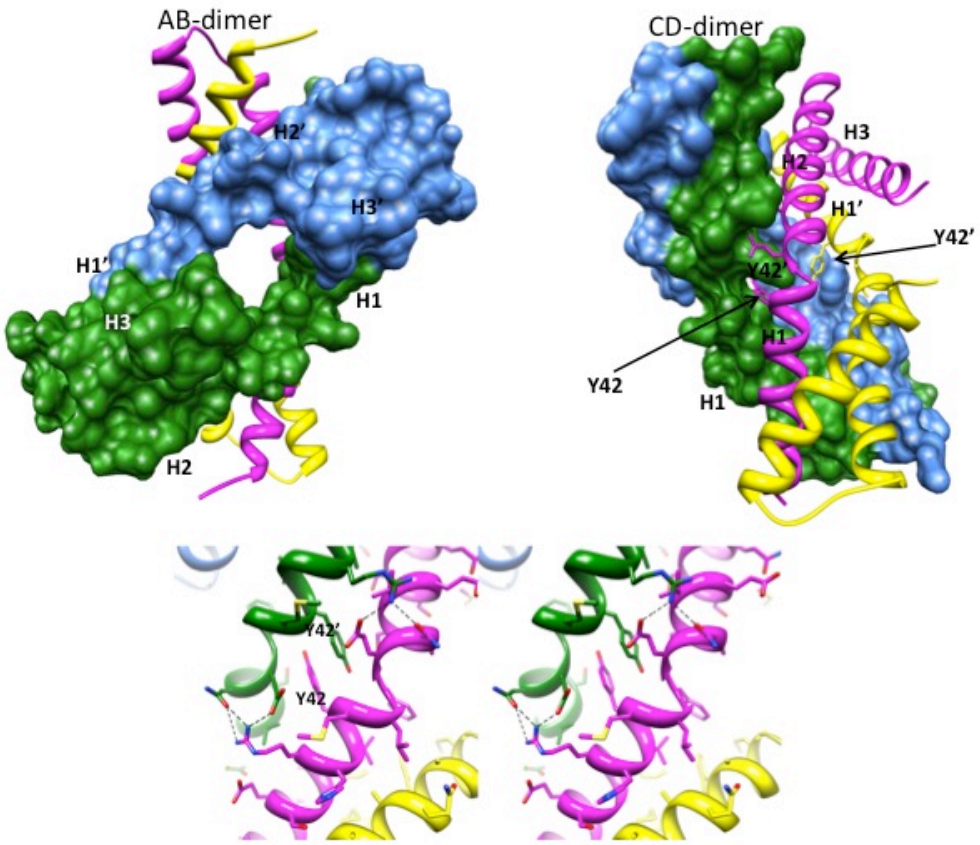


Figure 2.8 Tetrameric assembly of RI as a dimer of dimers. Second dimer is represented by filled model for clarity. Below is stereo close up of aromatic stacking interactions between the dimers.

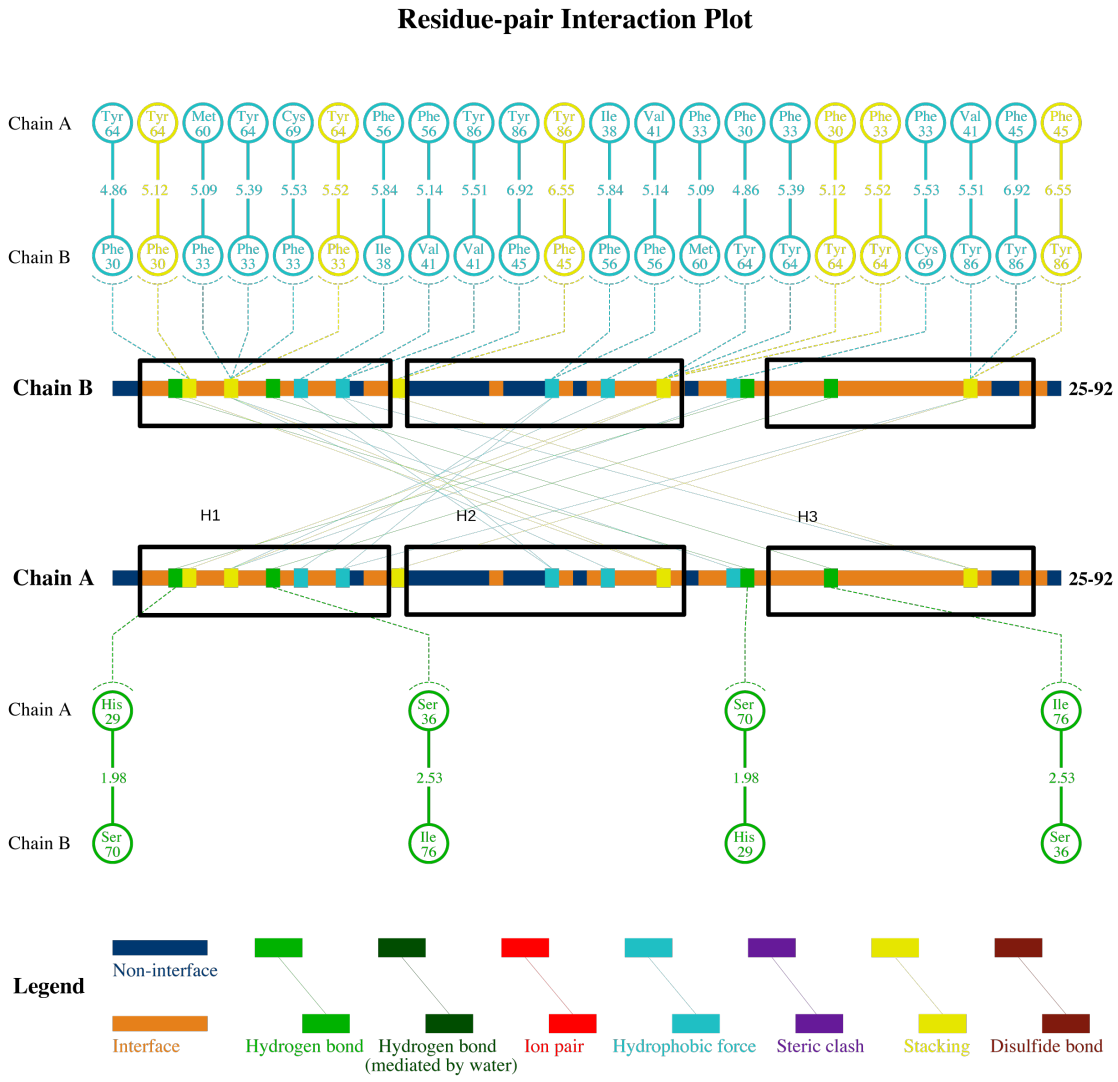


Figure 2.9 RI AB dimer interactions. 2D representation of all non-bonding interactions between RI subunits of dimer AB (GraLab).

three helices architecture of the RI and their oligomerization propensity might play a role in regulating its binding to T, the anti holin (*vide infra*). The structure of RI by itself leads us to an important question, which oligomeric form of RI can bind to T4 holin T.

sRI-sT complex

LIN requires the interactions of the periplasmic domains of the holin T and the antiholin RI. We have used the truncated proteins for the complex formation, without their N-terminal trans membrane regions. Even the truncated T showed a tendency of polymerization and aggregation, similar to full length T. The aggregation of T was prevented, by eluting the protein directly in to two-fold molar excess of purified RI, through the RI-T complex formation. The complex formation was further confirmed by gel filtration purification, which indicated the presence of RI-T complex with 2:2 stoichiometry. It has been shown previously that the periplasmic domain of RI alone is sufficient to bind to T and inhibit the holin polymerization. It has also been shown that the truncated RI should bind to the periplasmic region of the T. In order to understand the structural basis of the holin inhibition mechanism by anti holin (RI), we have crystallized the complex between the truncated proteins of RI (residues 25-97, sRI) and T (residues 56-218, sT hereafter). The x-ray crystal structure of sRI-sT was solved using the SAD method with crystals of selenomethionylated, complexed proteins formed in the space group $P6_522$ at 2.6 Å resolution (Table 2.2). One single complex of sRI-sT crystallized in the asymmetric unit (see Figure 2.10 for the nomenclature), however a thermodynamically stable hetero tetramer was clearly visible in the crystal lattice.

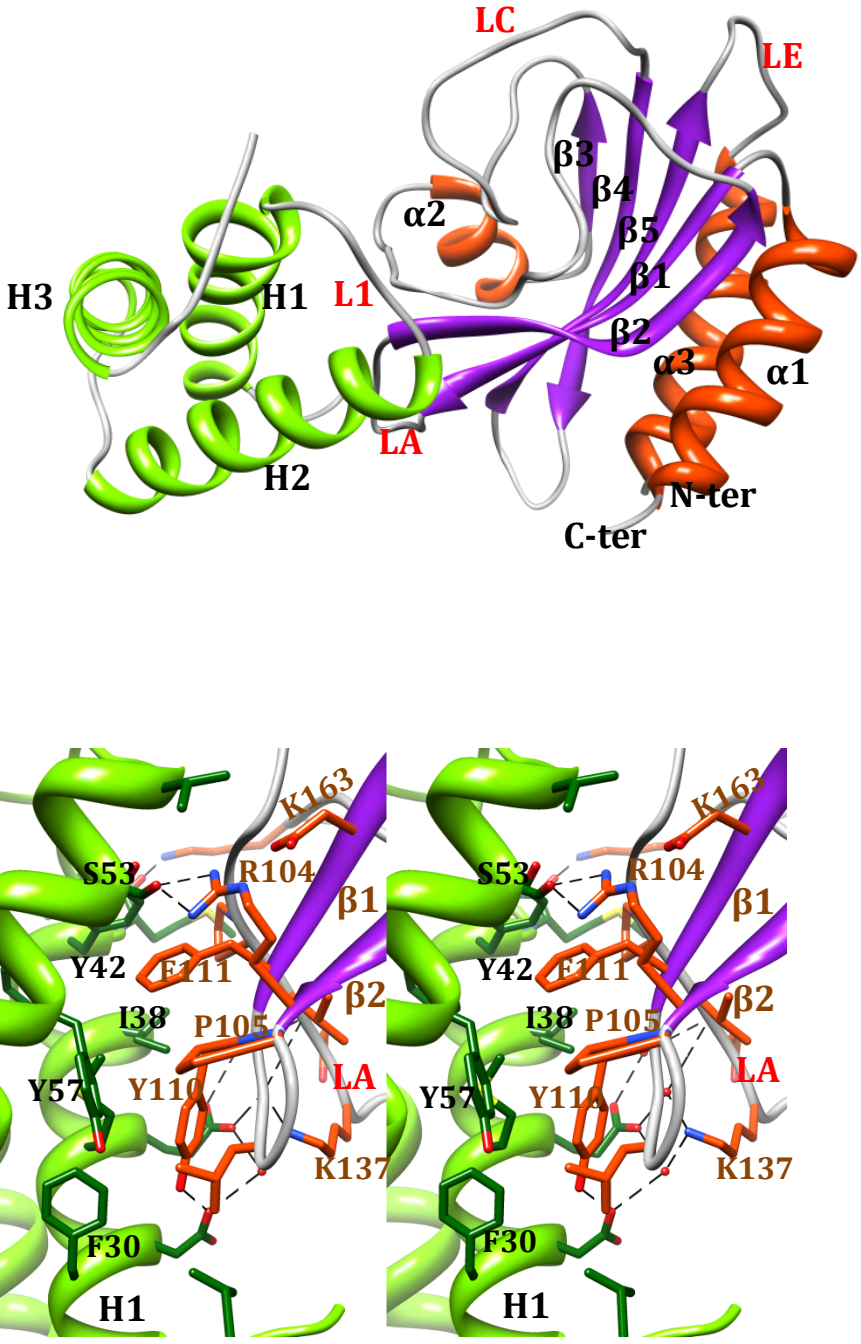


Figure 2.10 sRI-sT dimer. Below is stereo close up of the protein-protein interactions.

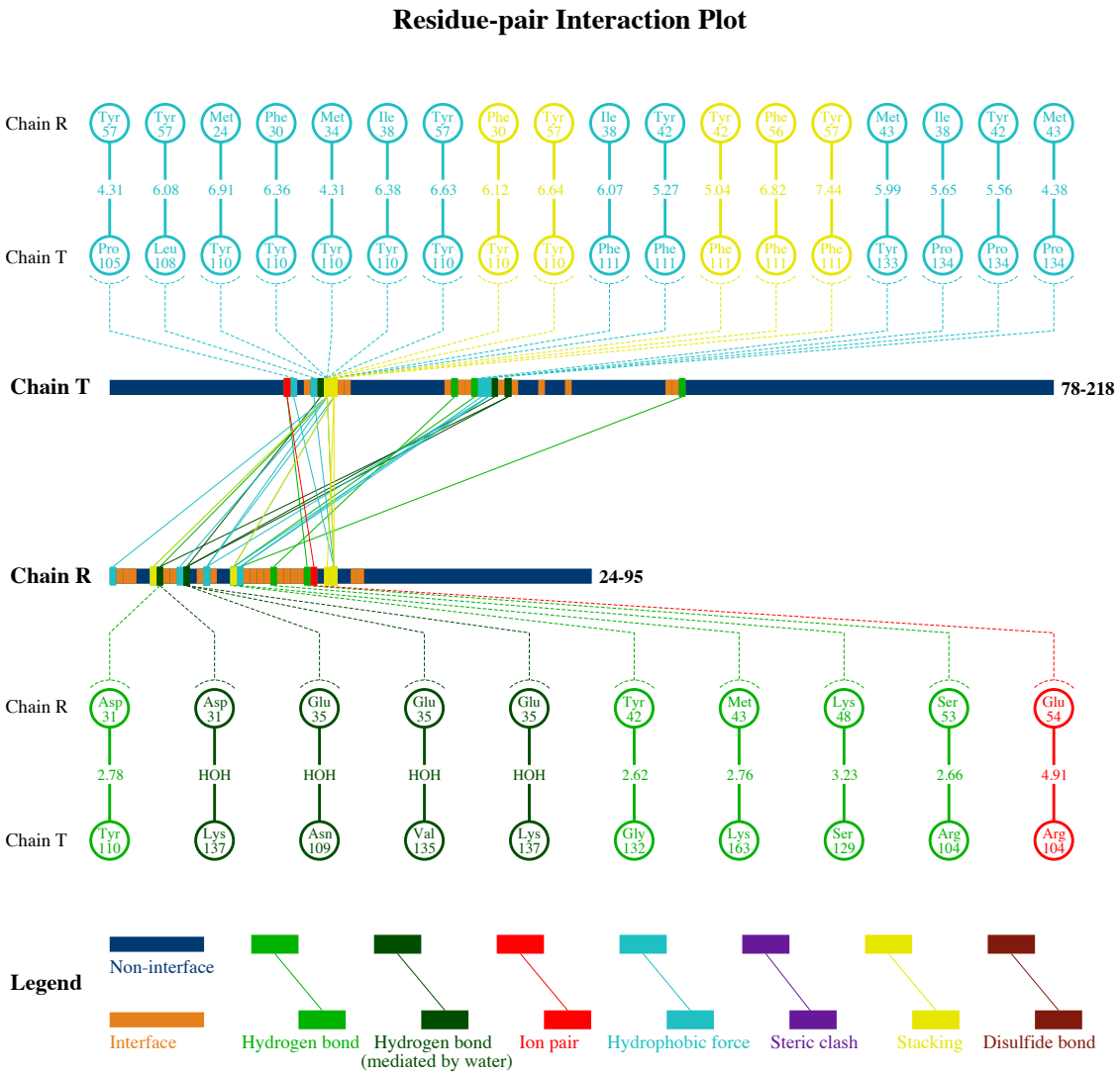


Figure 2.11 sRI-sT interactions. 2D view of non-bonding interactions between sRI and sT including hydrogen bonds mediated by water (dark green).

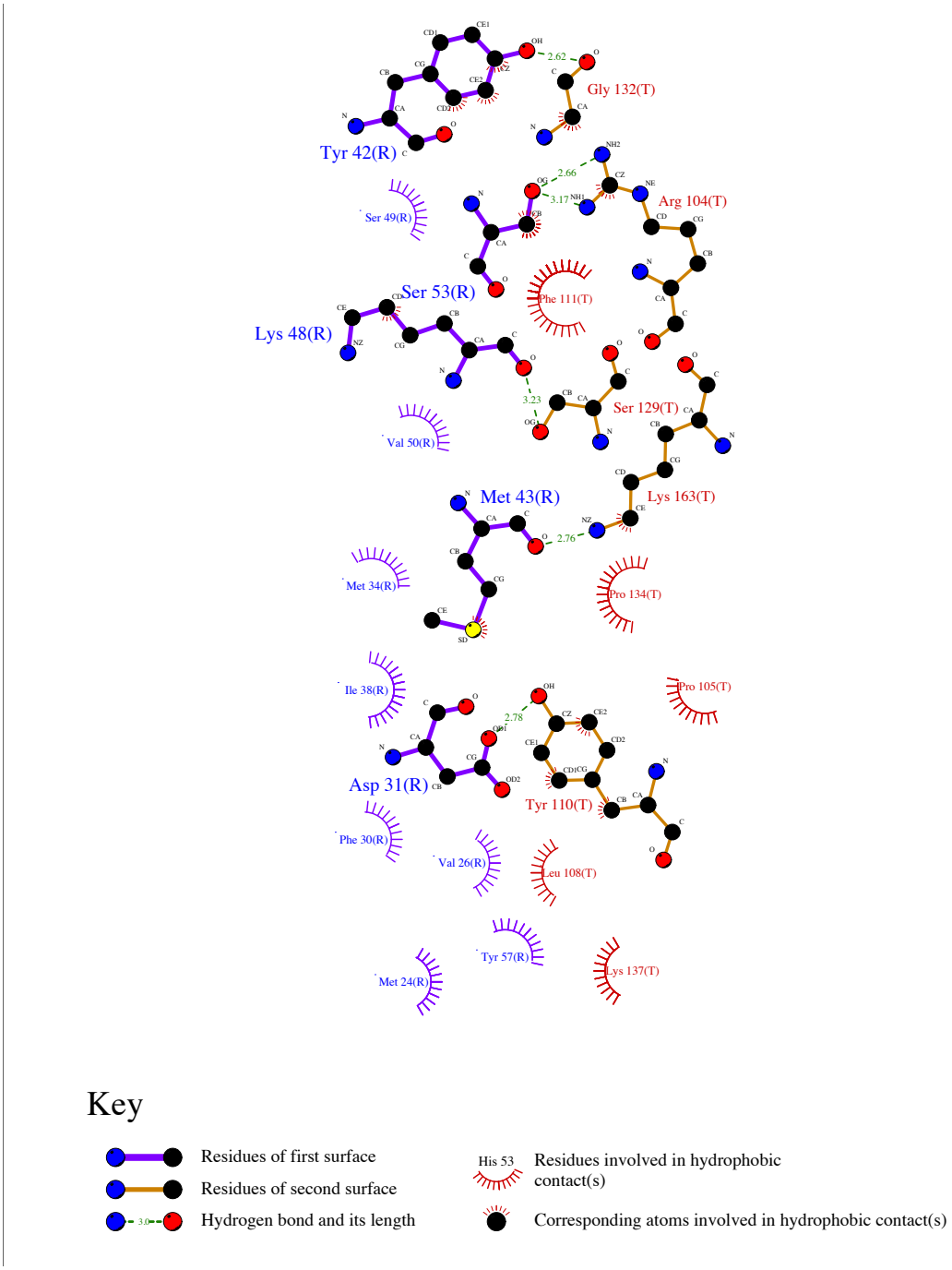


Figure 2.12 Ligplot of sRI and sT interactions within the dimer.

The final model was refined with the R factor of 19.8% and R_{free} of 23.8%, contains residues 78-218 of sT and residues 26-95 of sRI protomers. Residues 56-77 of sT and 24, 25, 96, and 97 of sRI are not observed in the electron density map and therefore omitted from the structure. 2D display of the binding region between RI and T is shown in Figure 2.11 and Figure 2.12. Two subunits of T are in close proximity with the two subunits of RI as shown in the figure on page 51 and discussed in later paragraphs.

RI binds to T as a monomer with its tail swapped in to the head

In contrast to the dimer-dimer assembly of the apo structure, RI molecules in the RI-T complex structure exist as monomers. Compared to the tetrameric apo RI subunits, the H1 helix of the holo monomeric RI structure undergoes a dramatic movement accommodated with some conformational changes in the H1-H2 kink region. In the RI holo structure, the helices H2 and H3 (the head region of RI), the disulfide bond between the residues Cys69 and Cys75 of this region and the H2-H3 connecting loop all remain mostly unchanged upon RI-T complex formation (Figure 2.13). In the RI apo dimer structure, the head region of one subunit formed a binding groove between the helices H2 and H3 for the accommodation of the H1 tail from the next subunit. In a striking contrast, in the holo monomeric structure of RI, the H2-H3 groove of the head region is now accommodated with its own H1 tail. The H1 helix of the holo structure undergoes a dramatic positional change (maximum displacement of 15 Å) to swing into the H2-H3 loop region. The H1 tail of the symmetry related subunit of apo RI dimer is now

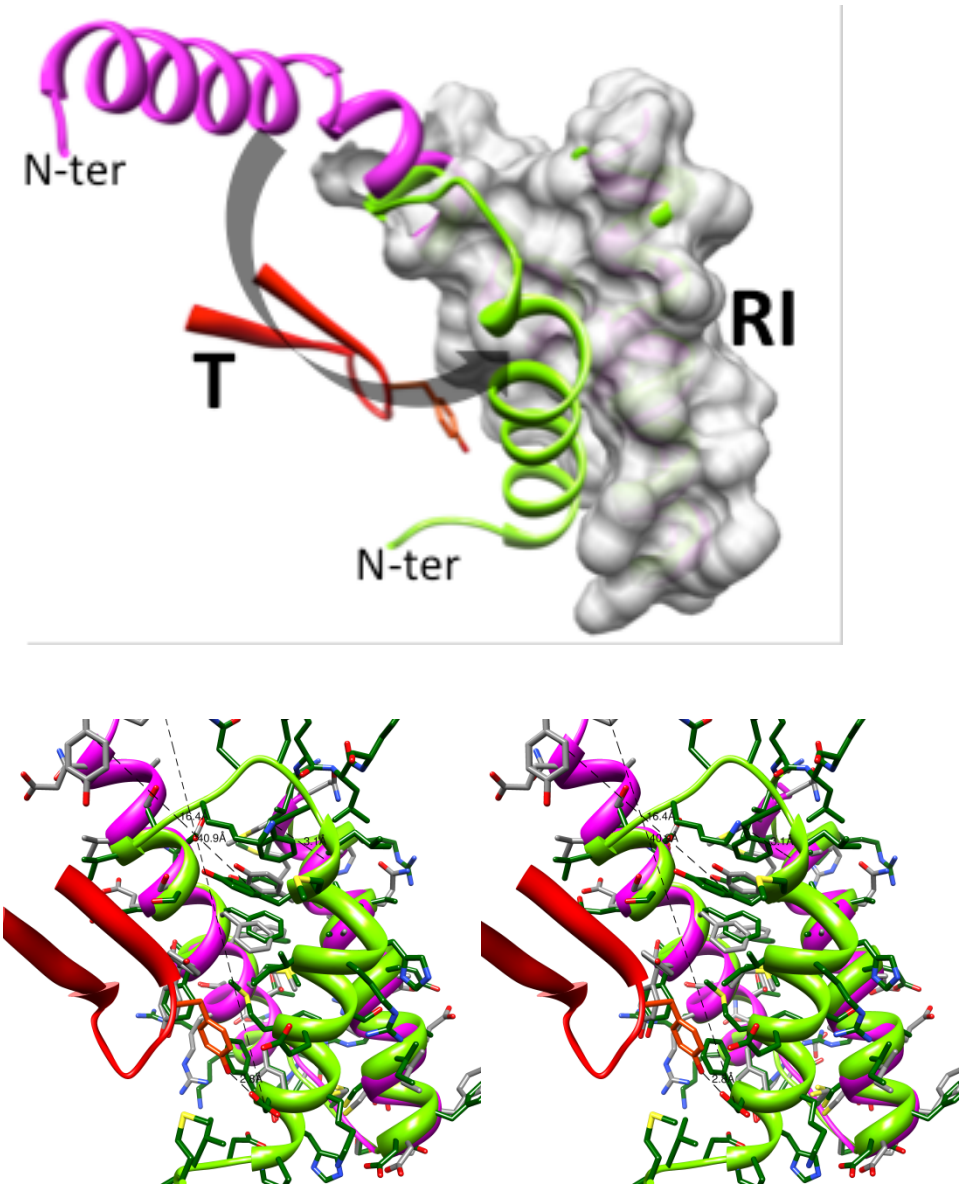


Figure 2.13 Conformational changes in RI. Helix H1 of sRI swoops from the open form (RI dimer) shown in purple to the compact monomeric form in green. Below is stereo close up.

swapped with its own tail region in the holo RI monomers as shown in Figure 2.13. The overall conformation of H1 helix of the holo structure remains very similar compared to that of the apo RI structure except some local variations in the side chain rotomer conformations of residues Met34, Phe45, Phe56, Tyr57 and Tyr86 (Figure 2.13). However, the kink region of the holo RI structure near the Phe45 residue undergoes significant conformational change to become a loop structure and augments the dramatic movement of the tail region. Compared to the apo RI structure, the disulfide bond locked H2-loop-H3 head region of the holo RI monomer remains unchanged and its N-terminal H1 helix tail swings into its own head (Figure 2.14).

The next important question is that what induces the dramatic movement of this H1 helix and monomeric conformation of RI? Is it possible that the monomeric form of RI is induced by its partner protein T? Based on the detailed analysis of RI-T interactions (see next paragraph), it appears that the inter protomer contact points do not seem to be capable of inducing such a big change in the RI structure. Obviously, if the unbound RI can exist only in the tetramer or dimer form, then the oligomer needs to dissociate into either the tail extended or compact monomers. It is most likely that the free RI molecules can exist as monomers, dimers and tetramers in solution, in equilibrium with one another. However, only the monomeric form of RI is capable of complex formation with T. In physiological conditions, the level of overexpression of RI and its total concentration in the periplasm should control this equilibrium.

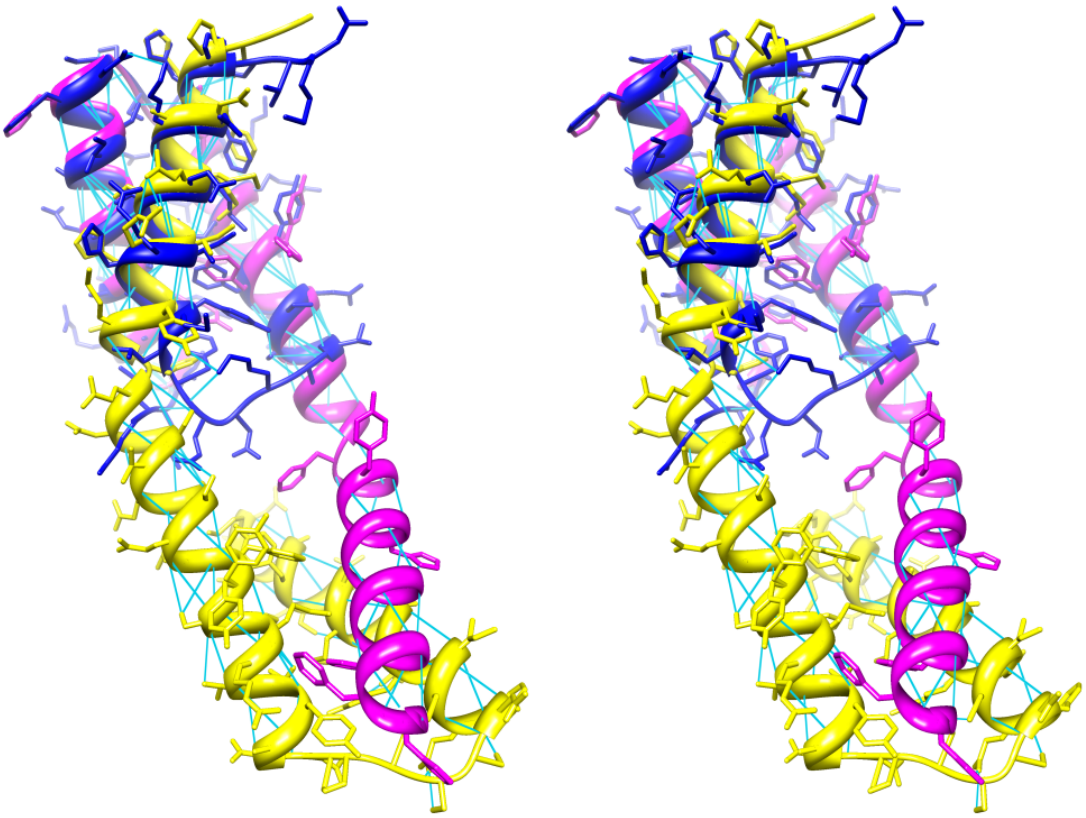


Figure 2.14 Stereo overlay of two conformations of sRI. The compact monomeric form that binds sT is in blue. Helix H1 has supporting interactions with its own subunit in the monomeric form.

The structure of sT

The structural fold of the holin was also a complete mystery as its amino acid sequence is unique and had no significant match with any protein data bank structures. The sRI-sT complex crystal structure presented here also provides the first detailed molecular description of the sT protein. The sT structure is composed of a central, twisted, antiparallel five-stranded β-sheet formed by β1 - β5 strands, two α helices on one side of this sheet (a N-terminal α1 helix and a C-terminal α3 helix) and on the opposite side is a mixture of loops and a short α-helix, α2. The small loop between β1 and β2 is labeled LA, β4 and β5 is LD and β5 and α3 is LE. The larger loop between β2 and α2 is designated LB (residues 119-140), and β3- β4 connecting loop is LC (residues 143-170). An intra subunit disulfide bond links the Cys175 of β4 strand with the Cys 205 of the C-terminal helix.

The sT crystal structure was searched against the protein data bank with VAST and DALI servers. Despite a lack of sequence homology to any previously determined structure, the sT structure revealed a number of unexpected similarities to several GAF domains in its three-dimensional structure (Figure 2.15). The GAF domains (acronym derived from first three letters of cGMP-dependent phosphodiesterases (PDEs), *Anabaena* adenylyl cyclases, and *E. Coli* FhlA proteins) have been discovered in 7400 different proteins found in all three kingdoms of life. The GAF domains have been shown to provide a variety of functions including binding of small molecules, protein-protein interactions (mostly dimerization) in hundreds of signaling and sensory proteins.

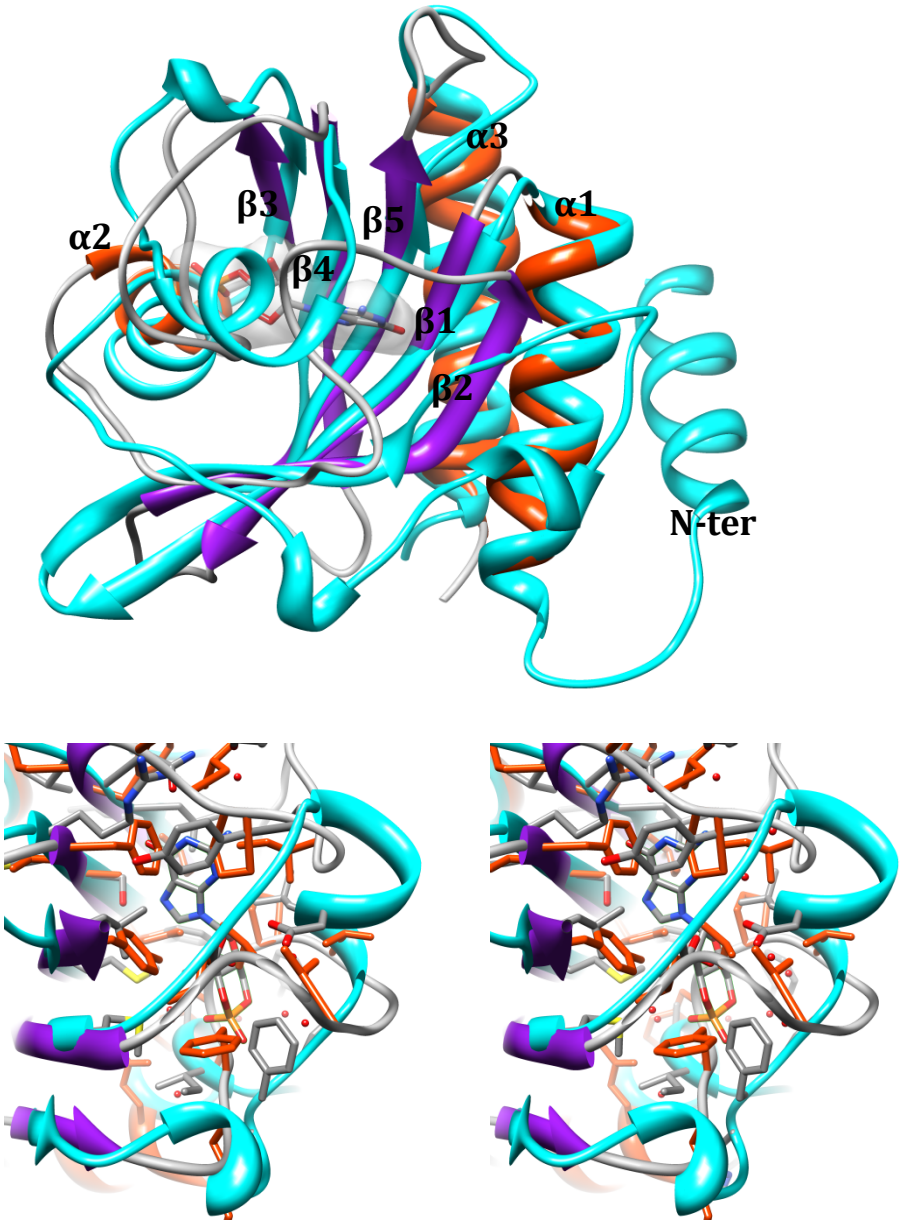


Figure 2.15 GAF domain of T. sT is superimposed on GAF domain (light blue) of cGMP binding protein (PDB: 2ool). cGMP molecule is represented by light grey surface volume.

GAF fold is also similar to PAS domains, which are too involved in sensory and signaling pathways.

The common concept among GAF/PAS domains is that they can interact with diverse set of small molecule regulators. The putative ligands for many of the GAF/PAS containing proteins remain subtle, suggesting that a number of small molecule signaling and sensory pathways are yet to be explored. Based on several previously published GAF domain crystal structures, that hold the highest structural similarity to sT, the sT structure should possess a small molecule-binding pocket between the central beta sheet and the loops LB-α2-LC region. This is rather surprising information with respect to the holin family of proteins. Closer analysis of this region (Figure 2.15) suggested that the loops LB- α2-LC can rearrange themselves to accommodate small molecules like cGMP and residues around Glu141, His 145, Asp 136, Phe159, and Tyr191 may play critical role in the binding. However, in the current RI inhibited conformation, this region is completely buried and not accessible for any small molecule binding. But it seems that this region is capable of forming a small molecule-binding pocket by rearranging the LB and LC loops, in absence of RI. However, we are not aware of any possibilities of small molecule binding to holin proteins. Given the diverse functional background within which GAF domains were found, it's not expected that they all possess identical functions. However, for now, we conclude that holin T utilizes GAF fold, an evolutionarily adopted signal sensor structure, for its signal transduction function. The role of small molecules in the sequence of events of phage lysis in context to GAF like holin warrants further investigation.

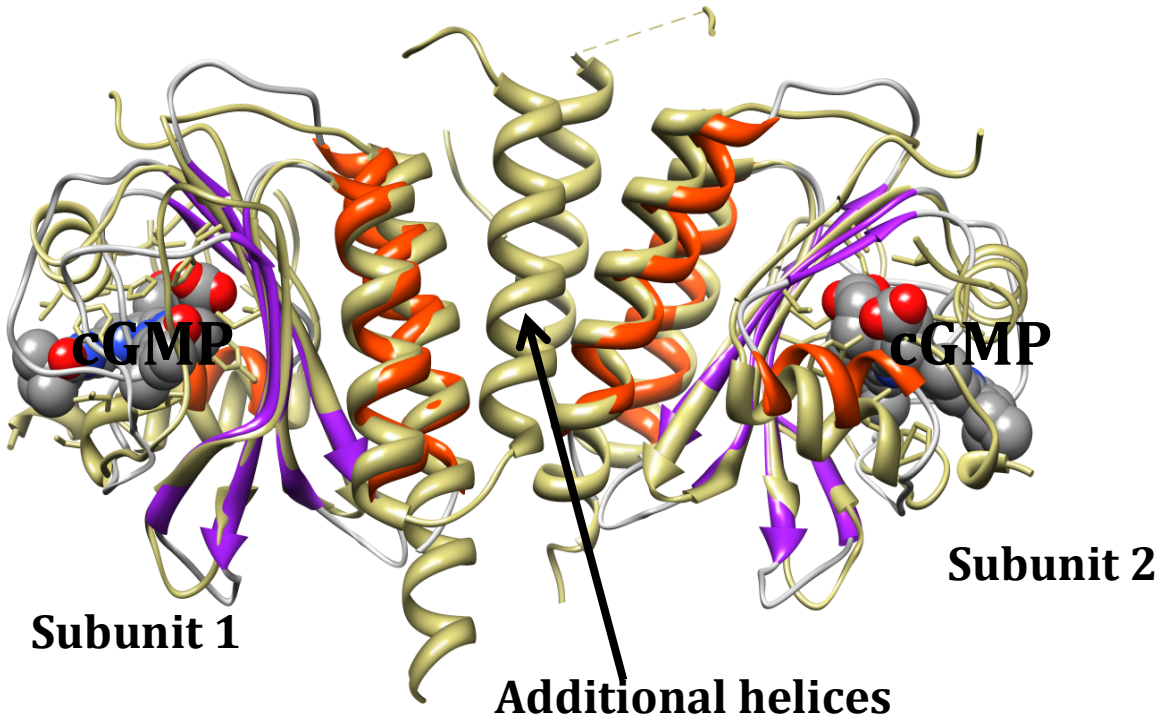


Figure 2.16 Model for holin dimerization. Dimerization could proceed via their α1 and α3 helices. Two holins (indicated by Subunit 1 and 2) are superimposed on top of GAF dimer (PDB: 2ool).

Another important aspect of GAF protein is that it has been shown to regulate protein function through dimerization. In this context, based on structural similarity, the sT proteins can oligomerize through their α1 and α3 helices. Oligomerization and polymerization is the most important aspect of the holin family of proteins, as they need to form the lesions in the bacterial cell membranes during the process of phage lysis. Figure 2.16 shows the superposition of the sT subdomain of the sRI-sT complex structure on top of the GAF dimer (PDB ID: 2ool). In GAF domain, an additional helix is also involved in the dimer formation (indicated by an arrow in Figure 2.16), a 3+3 helix bundle is formed in the dimer interface. The residues 56-77 at the N-teminal of the sT subunit are disordered in the crystal structure and the protein was over expressed without its transmembrane region. It is likely that, in the absence of anti holin, the holin initially forms dimers using a 4 or 6 bundle helices at the interface, similar to many other GAF proteins. Interestingly, the sT-sRI complex structure also forms a tetrameric assembly as detailed in the next paragraph. However, in the current complex structure, the α1-α3 helix region of two sT subunits are moved apart from each other by the binding of two RI molecules on either side of the sT-sT contact points (Figure 2.17 and next paragraph).

Tetrameric assembly of the sRI-sT complex

All four subunits are required for the complex formation; the N-terminals of sRI and loops LA and LD of sT together form the interface of this hetero complex (Figure 2.17). This tetrameric assembly (2 sRI and 2 sT) should not be confused for a commonly

referred hetero (dimer-dimer), as all four protomers are required for the oligomerization. In absence of sRI, the current proximity of sT subunits cannot remain as a stable dimer as only small portions of the protomers interact; so is sRI. However, within the tetramer, each monomers of sRI forms several non- bonding interactions with both sT molecules, on either side of the sT-sT contact points. For the sake of simplicity, we will describe the inter protomer interaction on one side of the sT-sT contact point with respect to one subunit of sRI; identical interactions are observed on the opposite side of this interface. Figure 2.18 depicts the close up of the interacting region between sRI and sT. Each sRI molecule within the 2sRI-2sT tetrameric assembly, interacts with both the subunits of sT through the network of nonbonding interactions. As shown in Figure 2.17, residues of all three sRI helices (the H1-H2-H3 of the compact RI structure, *vide ante*) are involved in the nonbonding interactions with the β1-LA-β2 region and LB loop region of the T subunit. The residues Arg104-Phe111 of the sT subunits near the LA loop make many interactions, which are critical for the complex formation. The residue Arg104 of β2 of sT forms two hydrogen bonds with Ser53 of sRI subunit. The side chain atoms of residue Phe105 of sT pack directly into the side chain atoms of Tyr57 of sRI subunit (the CG tip of sT-Phe105 is at 4.1-4.2Å distance from the phenyl carbon atoms of sRI-Tyr57). The residue sT-Leu108 makes hydrophobic interactions with the side chain atoms of sRI-Val26 and sRI-Tyr57. The residue sT-Tyr 110 hydrogen bonds with the side chain OD1 oxygen atom of the sRI-Asp31 and makes hydrophobic interactions with residues sRI-(Phe30, Glu35, Ile38 and Met34). The residue sT-Phe111 points into the

hydrophobic pocket formed by residues sRI (Tyr42, I38, S53, Phe56 and Met34) at less than 4.5 Å distance. The sT-LB loop adjacent to the β2 strand of sT also makes

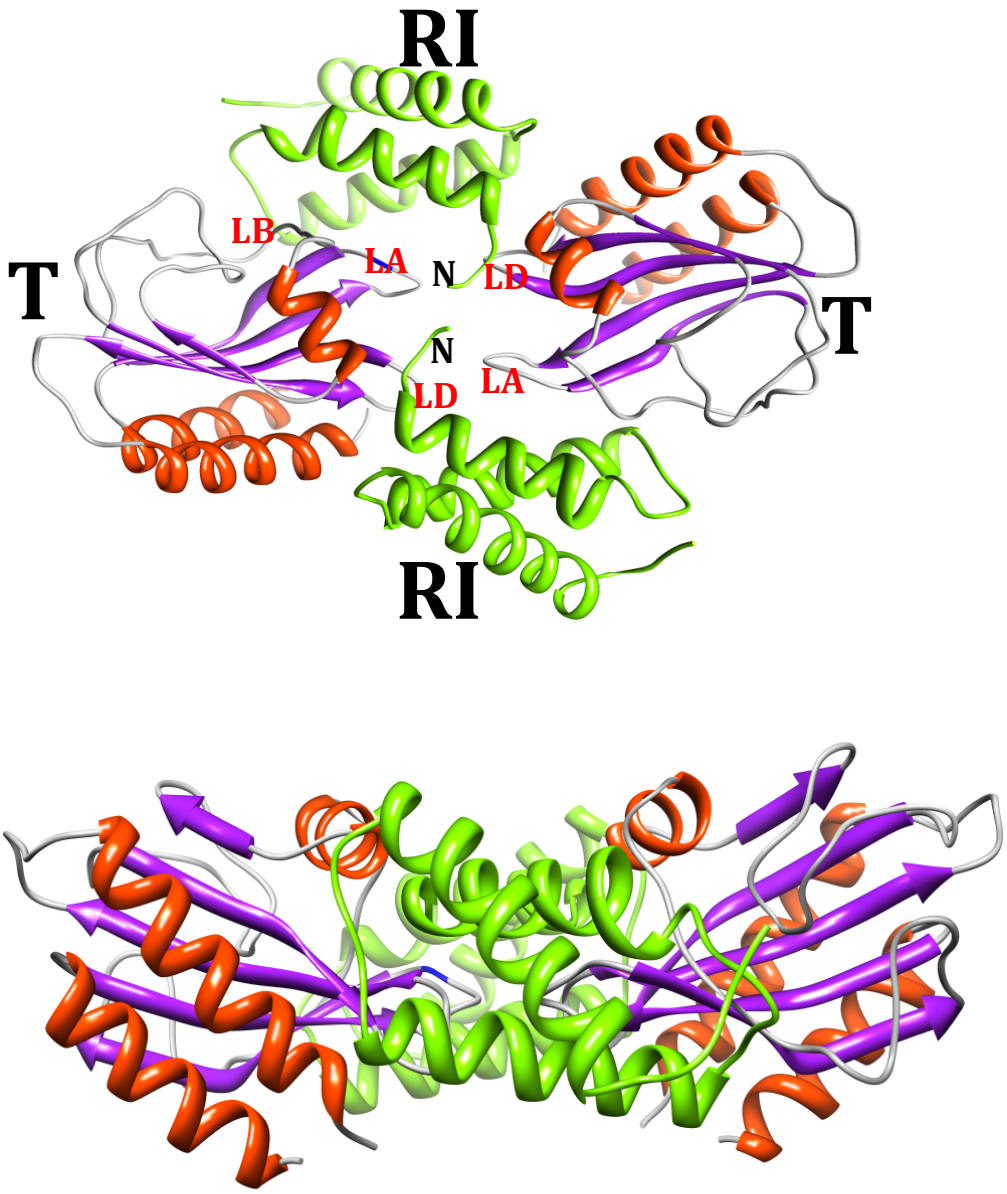


Figure 2.17 Tetrameric assembly of the holin-antiholin complex.

contact with the sRI helices. The residue sT-Pro134 of loop LB is in close proximity with the side chain atoms of residues sRI (Ile38, Tyr42 and Met43) to favor hydrophobic interactions. The main chain carbonyl atom of sT-Asp136 forms two water mediated hydrogen bond interactions with sRI-Glu35. The residue sT-Lys137 forms water mediated hydrogen bond interactions with the side chain atoms of sRI-Asp31 and sRI-Glu35. The sT-Lys163 residue of the loop LC hydrogen bonds with the main chain carbonyl oxygen atom of sRI-Met43. All these interactions are depicted in Figure 2.18.

The sRI subunit also interacts with the second subunit of sT. The residue sRI-Asp67 near the end of the helix sRI-H2 salt bridges with the side chain NH1 atom of the sT-Arg 212 of the α3 helix. The NZ atom of residue sRI-Lys65 hydrogen bonds with the main chain carbonyl oxygen atoms of sT-Arg216 (3.1 Å distance) and sT-Asn179 (3.1 Å distance). Residues Phe178-Tyr184 of the LD loop region of sT subunit also make many nonbinding interactions with the sRI subunit. The residues phe178, Leu180, Asn182, Ile183 and Tyr184 of sT subunit make hydrophobic interactions with the side chain atoms of residues Val26, Asp27, Phe30 and Asn25. Also sT-Asp181 of loop LD forms hydrogen bonds with the hydroxyl group of sT-Tyr57 (at 3.2 Å distance) and NH2 atom of sT-Arg61 (at 3.2 Å distance).

In addition to these interactions, the chain carbon atoms Asp181 and Ile183 of one subunit of sT protomer contact the Leu108 of the second sT subunit. The residue sT-Asn107 of loop LA near this contact point is disordered. Therefore, within the sT-sRI tetrameric assembly, one subunit of sT contacts the other subunit through the tip of the loops LA-LD' and LD-LA'. These are the only interacting points between the sT

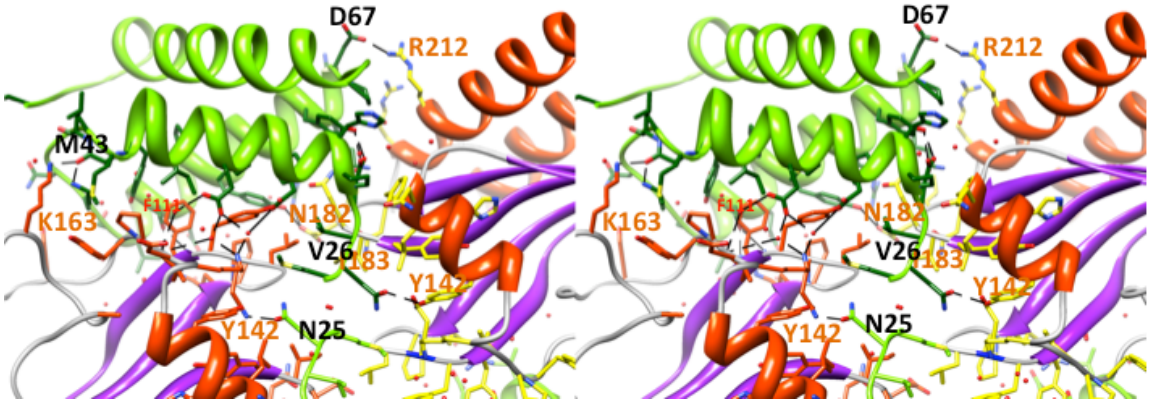


Figure 2.18 RI in tetrameric assembly. Stereo close up of antiholin interactions within the complex. Each sRI interacts with two holins.

subunits. However, each sRI molecule connects both the sT subunits through a network of nonbonding interactions, on both sides of the sT-sT contact points to form a stable tetramer. As a result, the anti holin RI moves the α1 and α3 helices of two holin subunits away from each other by a distance of 28 Å, possibly preventing holin polymerization. In absence of anti holin, the α1 and α3 helices of each holin subunit might interact to dimerize and then polymerize. The complete mechanism of polymerization needs to be explored further.

sRI tetramer is capable of binding DNA

Finally, the tetrameric assembly of RI, discussed above, satisfies major criteria of a DNA-binding protein: it has a classical helix-turn-helix motif, it satisfies various spacial and charge complementarity requirements, and it could fit in the major grooves of DNA. The tetramer is predicted to be able to bind a piece of DNA that is at least 30 bp long. Figure 2.19 displays computer-docking models of two possible ways the tetramer could bind DNA.

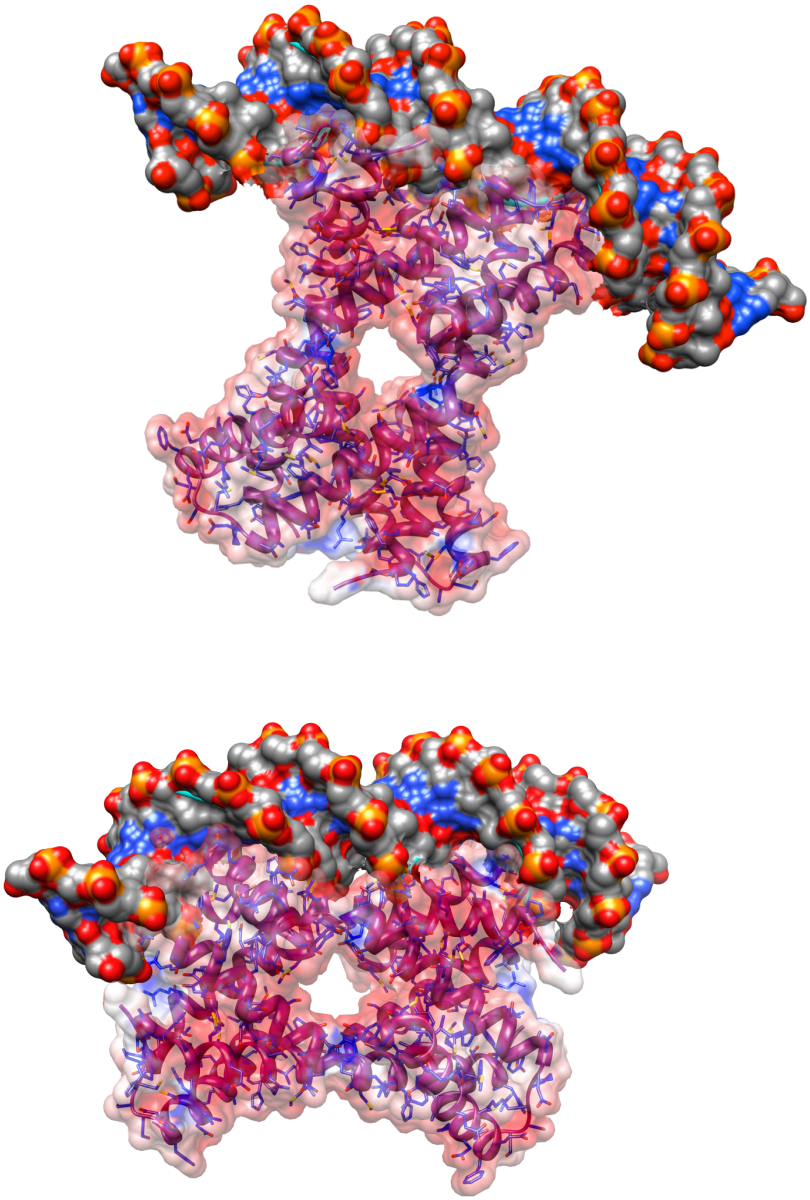


Figure 2.19 DNA binding models. Computer docking of DNA binding by sRI tetramer. The top model has a higher predicted probability.

Detection of macromolecular assemblies with EBI-MSD service PISA

By definition, a macromolecular assembly is a complex of more than one polypeptide (or nucleotide) chains that is stable in the native environment. The way the chains assemble represents the protein quaternary structure (PQS). Often times, protein quaternary structure can be a Biological Unit that performs, in its quaternary state, a certain biological function.

It is well known that the physiological function of a macromolecular assembly is closely related to its 3D structure. Determination of PQS is not trivial. Different experimental techniques have been developed to study various properties of macromolecular assemblies, but deducing a 3D structure proved to be difficult (29). Scattering techniques such as neutron and X-ray could deliver information on the chemical composition and possibly multimeric state of the assembly. Dynamic light scattering could help to guess a 3D shape from mobility measurements. Electron microscopy offers relatively poor resolution and is only suitable to study large complexes. Finally, NMR is capable of getting atomic coordinates of macromolecular complexes but has limitations on the size of a complex and is not ideal for studying protein assemblies.

More than 85% of PDB entries are structures solved by means of X-ray diffraction on crystals (30). Crystallography is a very special technique in that a protein crystal is made up of assemblies. It is expected that PQS makes building blocks for the crystal. There is no need to dock subunits to try ascertaining a correct oligomeric assembly; the docking is given by crystal structure. Macromolecular interfaces should be

viewed as an additional and very important aspect of protein crystallography. The problem is to determine a correct oligomeric assembly from a PDB entry. A PDB file contains atomic coordinates of the Asymmetric Unit (ASU). All space symmetry group mates of ASU comprise a Unit Cell. Generally speaking, neither ASU nor a Unit Cell has any direct relation to the protein quaternary structure. The PQS may be made of a single ASU, a part of ASU, several ASU, or even several parts of ASU.

Stability of protein complexes depends on properties of protein-protein interfaces that include free energy of formation ΔG_{int}, solvation energy gain ΔG_s, the interface area, hydrogen bonds and salt bridges across the interface, and hydrophobic specificity (31). One could view a protein crystal as a packing of assemblies that are separated by biologically insignificant contacts, while protein assemblies are made up of monomeric units with biologically relevant interfaces between them. The real problem is to find a way to separate biologically relevant interfaces from superficial ones caused by inter-assembly and crystal packing.

Over the years a considerable effort has been made in attempts to assess the significance of the interface from its properties. Most often used discrimination criterion is the interface area, but this does not work for large proteins. Free energy gain of interface formation does not consider if energy measure can be uniform for all weights and shapes. P-value of hydrophobic patches, which measures probability for the interface to be more hydrophobic than found, also does not offer a very high success rate of discrimination between monomers and oligomers. Based on empirical evidence, there is no ultimate discriminating parameter to identify biologically relevant protein

interfaces, even for the dimers (31). Formation of oligomeric complexes is probably a collective process that involves a set of interfaces. As such, significance of an interface should not be separate from the context of the protein complex.

PISA (*Protein Interfaces, Surfaces and Assemblies*) computational server is built on principles that it is not the properties of individual interfaces that really matter; it is the chemical stability of a protein complex (32). In PISA, macromolecular complexes are identified as chemically stable associations. Protein chains will most likely associate into largest complexes that remain stable. A protein complex is stable if its free energy of dissociation is positive:

$\Delta G_{diss} = -\Delta G_{int} - T\Delta S > 0,$

Where ΔG_{int} is the enthalpy of engaged interfaces (or the binding energy), T is the temperature, and ΔS is the cost in entropy of dissociation. Complexes with positive ΔG_{diss} are considered chemically stable.

A multimeric complex may dissociate in different ways and even have more than one pattern of dissociation. For example, a homotetramer may dissociate into two dimers or four monomers. PISA identifies the preferred dissociation pattern as one with the lowest ΔGdiss by analyzing all possible dissociation paths (30).

As mentioned earlier, the elongated structure of RI, containing one molecule of RI in ASU, was evaluated for probable macromolecular assemblies that would be stable in solution with PISA server. Analysis of protein interfaces of RI suggested a homotetramer (ΔG_{diss} =13.8 kcal/mol) dissociating into more stable homodimers

(ΔG_{diss} =26.6 kcal/mol). Macromolecular assemblies predicted by PISA were visualized and appeared to be in good agreement with experimental structures derived from X-ray diffraction.

Mechanism of action of antiholin RI resembles that of ecotin

Based on the two crystal structures presented above, RI mode of action bears striking similarities to the mode of action of another periplasmic protein ecotin.

Proteases are often kept from wreaking havoc by binding to proteins that are protease inhibitors. Ecotin is a periplasmic protease inhibitor in *E. coli* that binds to and blocks activity of virtually all serine proteases with the canonical Asp-His-Ser triad, regardless of the amino acid sequence or substrate specificity. This remarkable extent of specificity of inhibition classifies ecotin as a fold-specific inhibitor (33).

Ecotin is a 142 amino acid protein with a single disulfide bond that forms a stable homodimer of about 16 kDa per subunit (34). The monomer is a single domain made up of antiparallel seven-stranded β barrel and loops connecting the secondary structure elements. The striking feature of ecotin is its nonglobular shape, with ecotin dimer resembling a butterfly (35). The dimer is held together by antiparallel β ribbon interactions between the two C-termini in an arm-in-arm fashion. This extensive dimer interface characterizes ecotin as a dimer with swapped domains. The notion of domain swapping was first introduced by David Eisenberg (36). The term domain swapped means that a segment of one subunit extends a domain that packs into the second subunit. The critical element for function of ecotin is that 15 amino acids at the C-

terminal are packed with the symmetry related subunit (mate). Eisenberg proposed domain swap as a possible mechanism for evolution of multimeric proteins from monomers. Dimeric ecotin may have evolved from a monomeric inhibitor by domain swapping.

Ecotin forms a tetrameric complex consisting of two protease molecules and the ecotin homodimer in the E_2P_2 form (Figure 2.20) (37). Ecotin binds in a bidentate manner with two surface loop regions that are known as the primary and secondary sites. This makes each ecotin molecule to have three protein-protein interaction surfaces. It is important to note that each ecotin molecule contacts both of the protease molecules.

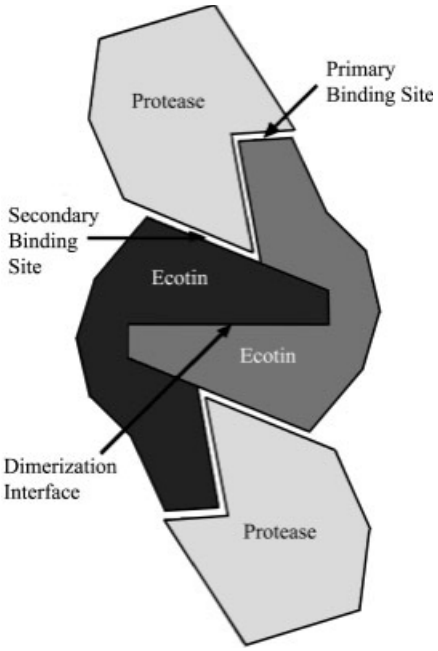


Figure 2.20 Tetramer of ecotin bound to two proteases. Ecotin forms a network complex (33).

The primary site of ecotin binds in a manner that resembles a protease substrate and forms up to eight main chain hydrogen bonds as a β sheet along the active site cleft. Ecotin is a very potent inhibitor due to the primary site loops, but the method of inhibition is still not well understood. The secondary site of ecotin is spatially independent. The secondary site binds to the protease at the terminal helix over 20 Å away from the active site and could form up to five additional hydrogen bonds. The binding at the secondary site accounts for the fold specificity of ecotin.

Ecotin tetramer could be thought of as a network (38). The advantage of a network over a chain is a potential to gain free energy of association through cooperativity of the interfaces. A network is potentially much stronger than the sum of its constituent parts, while a chain, such as a typical dimer of protease protease-inhibitor is only as strong as its weakest link.

All structural homologs of RI form homo oligomers

The apo RI structure was searched against the Protein Data Bank using the DALI server and no direct homologous structures were found. The top scoring structures were helical regions of other proteins or parts of helical bundles that only roughly resembled RI (Figure 2.21). However, analysis of the top 25 hits revealed important information about RI oligomerization. In spite of marginal similarity, all 25 structures formed some kind of higher order oligomeric assembly. Apart from DALI search, we also found two other homo oligomeric structures similar to apo RI in our literature search (39) (40). Interestingly, all these structures utilized their helical region for the homo

oligomerization. In general, it's natural that the free helices in solution can bundle together to form higher order oligomers. In case of RI, the elongated apo form should at least form a stable dimer in solution as suggested by the crystal structure and DALI search results. Therefore the RI should either exist as a higher order oligomer as shown in apo RI crystal structure or fold in to a compact monomeric form as observed in the RI-T crystal structure.

Modeling RI in the presence of its SAR domain

Based on the crystal structure of the periplasmic region of apo RI, it could exist as a dimer, tetramer and monomer. The structure is missing the N-terminal region that comprises the SAR domain. Within the lipid bilayer, the SAR domain should generally span as a single long helix. However, it has been shown in the crystal structure of the full-length lysozyme of coliphage 21 (R^{21}) that the SAR domain can undergo a conformational change. It folds into two anti-parallel alpha helices once extracted from the membrane bilayer, and packs against the C-terminal helical bundle of the enzyme (41). In order to visualize various positional possibilities, we modeled the SAR domain of the full length RI utilizing the crystallographic coordinates of the first 24 residues of R^{21} enzyme. Figure 2.22 depicts the modeled N-terminal SAR domain of RI subunits in a tetramer and dimer (shown as orange ribbon). Within a tetramer, the SAR domain has to stay at the periphery near the N-terminals of each subunit; suggesting a somewhat restricted movement. As shown in Figure 2.22, within a dimer, the SAR domain can either stay at the edge or fold towards the body (as indicated by the black arrow) of the

protomer. What is clear from the speculated model is that the oligomerization of RI is still possible, even in the presence of SAR domain. Figure 2.23 depicts possible positions of the RI SAR domain, in the RI-T complex structure.

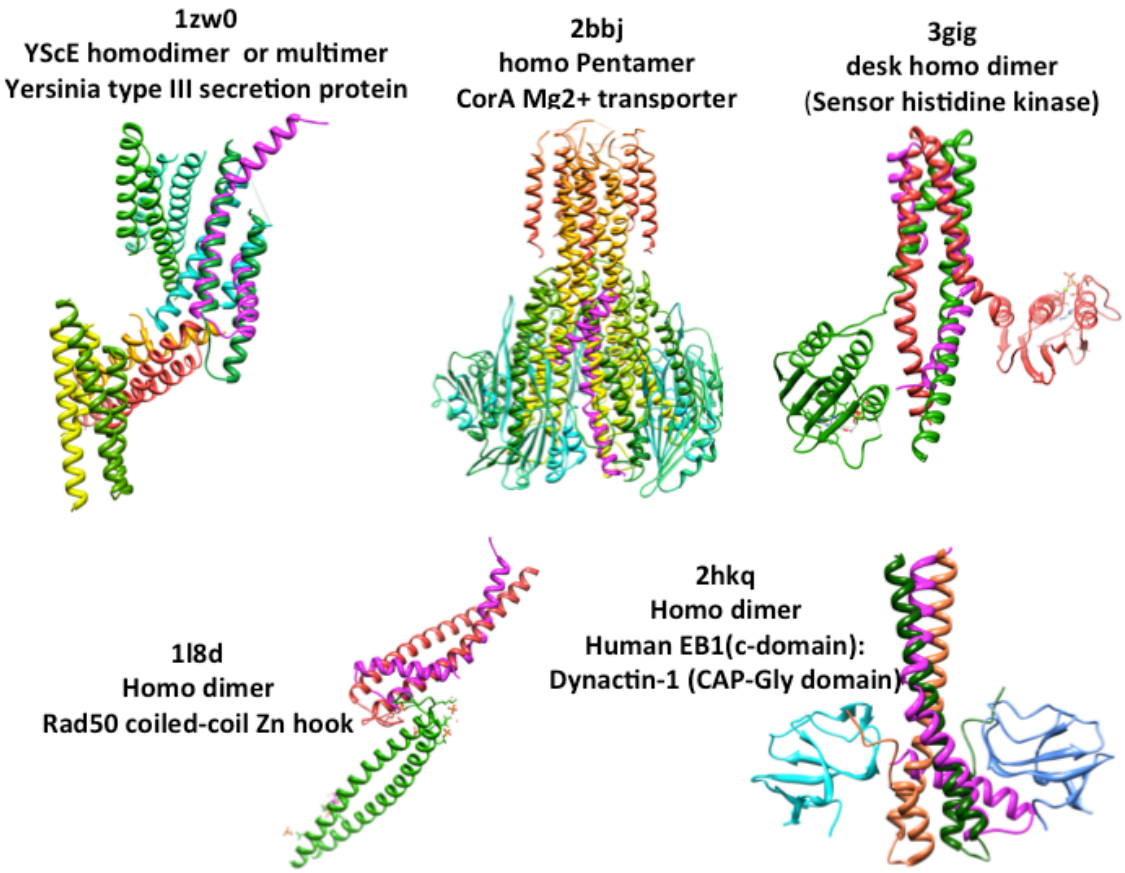


Figure 2.21 RI structural homologs. Top structural homologs of RI are homo oligomers. RI is superimposed on each of the top five structures (magenta helices). Search performed with DALI and VAST servers.

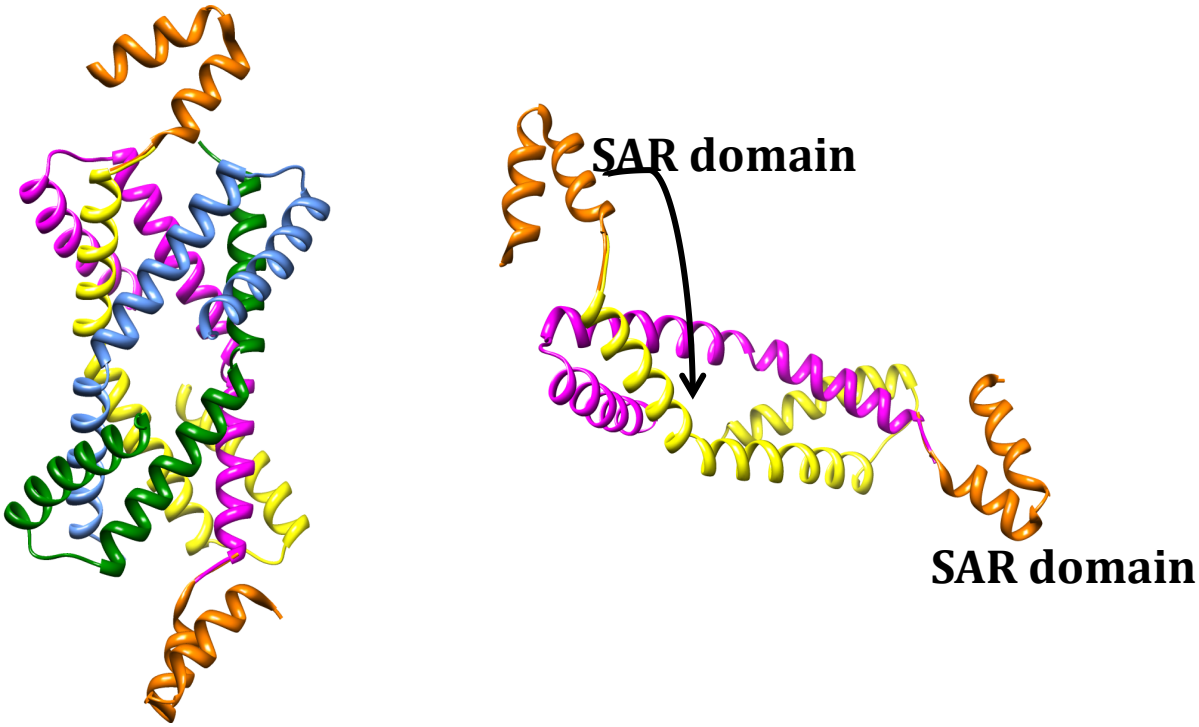


Figure 2.22 Models of the SAR domain in RI oligomers. RI tetramer (left) and dimer (right) with the modeled SAR domains (orange ribbons). Arrow indicates possible alternative packing for the SAR domain in RI homo dimer. For visual clarity, only two SAR domains are shown for the tetramer of RI.

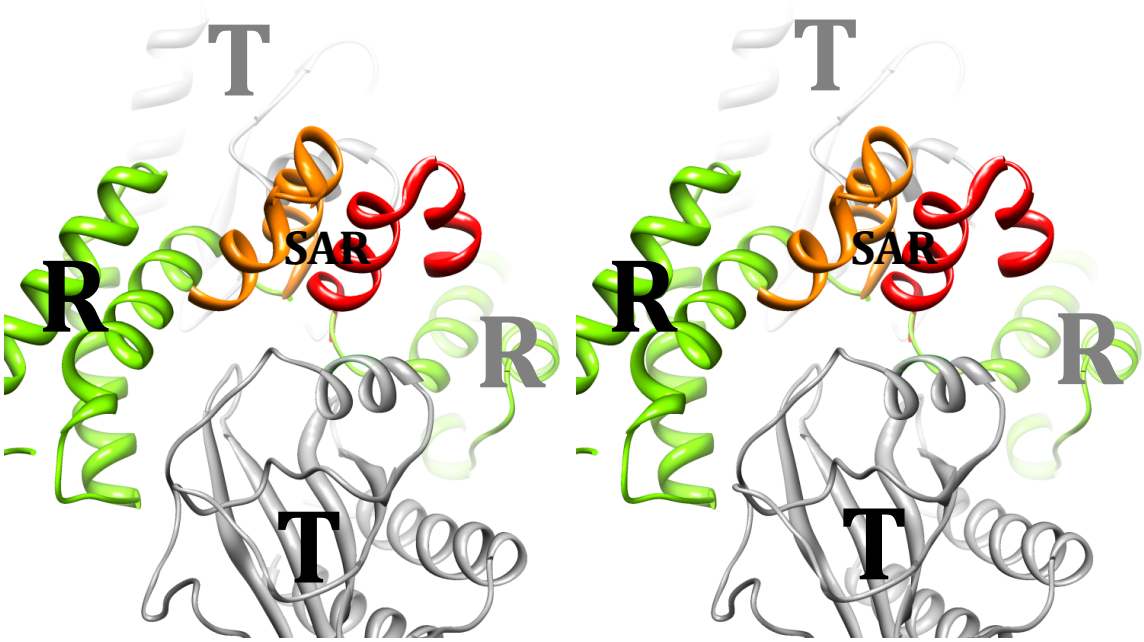


Figure 2.23 Model of the SAR domain of RI in RI-T tetrameric complex. Two SAR domains could form a helix bundle at the central contact points.

CHAPTER III

ANALYSIS OF Qβ A_2 AND OF *B. subtilis* MurA

BACKGROUND

Lytic bacteriophages are parasitic viruses that infect and replicate inside bacteria. Most phages accomplish host lysis using a muralitic enzyme, or endolysin, and a holin, which permeabilizes the membrane at a programmed time and thus controls the length of the vegetative cycle. By contrast, lytic single-stranded RNA and DNA phages accomplish lysis by producing a single lysis protein without muralitic activity (42).

The small, single-stranded RNA bacteriophage Qβ has a 4.2 kb nucleotide genome (Figure 3.1).

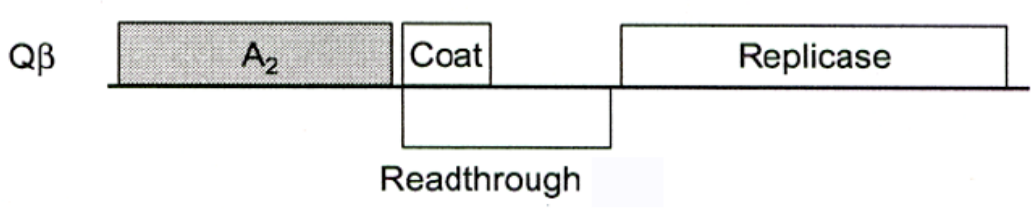


Figure 3.1 Qβ genome. Linear genome of ssRNA phage Qβ. The gene encoding the maturation protein A_2 (420 a.a.) is indicated by the shaded box.

The genome contains three cistrons, but encodes four proteins: maturation (A_2), coat, read-through coat, and replicase (43). Infective Qβ virions are composed of one genomic ssRNA molecule encapsulated by 180 copies of a combination of the coat protein and read-through protein, and a single copy of A_2 (44). Lacking the holin-endolysin machinery, Qβ produces only a single gene product to effect lysis. The maturation protein A_2, which binds the host sex pilus, has a second role as a lysis protein (43). It was observed that A_2 induction from a plasmid caused cell wall synthesis to cease at least 20 minutes prior to lysis (45). The host target of A_2 was identified as MurA, an enzyme that catalyzes the committed step of murein biosynthesis (45).

UDP-*N*-acetylglucosamine enolpyruvyltransferase (MurA) catalyzes the first committed step in the peptidoglycan biosynthesis (Figure 3.2) (46) by transferring the enolpyruvyl moiety of phosphoenolpyruvate (PEP) to the 3'-hydroxyl group of UDP-*N*-acetylglucosamine (UDP-NAG) (Figure 3.3). The product of this reaction is the 3-enolpyruvyl ether of enolpyruvyl-UDP-NAG. While most enzymatic reactions utilizing PEP as a substrate involve cleavage of the high-energy P—O bond, in this reaction the C—O bond of PEP is cleaved to transfer the enolpyruvyl moiety to a second substrate. The only other enzyme known to catalyze the transfer of the intact enolpyruvyl moiety of PEP to a substrate is 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) (47). MurA and EPSPS have a very similar structure despite an amino acid sequence identity of only 25%

The cell wall biosynthesis pathway is not present in humans. For that reason the pathway is an attractive pharmaceutical target. MurA is the only enzyme in the pathway

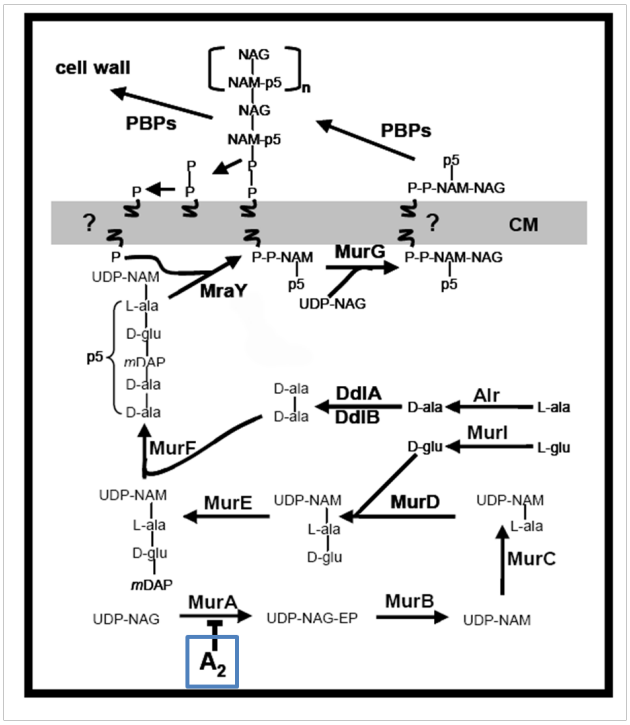


Figure 3.2 Cell wall synthesis pathway. A_2 (blue square) inhibits enzyme MurA.

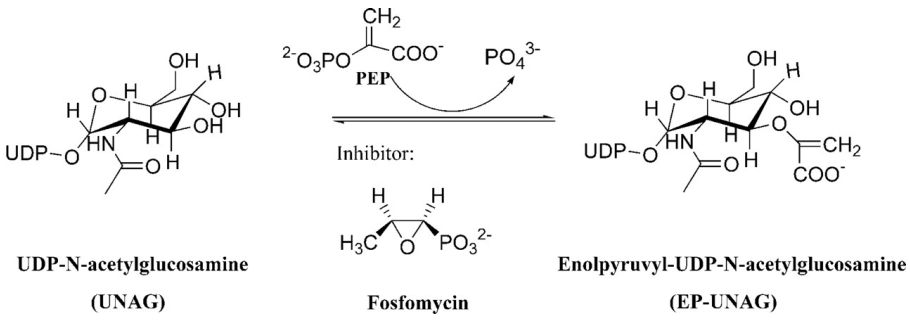


Figure 3.3 Reaction catalyzed by MurA.

that is inhibited by fosfomycin, a natural antibiotic (46).

From the MurA crystal structures it was concluded that unliganded MurA exists in an open conformation (49) (PDB: 1naw) whereas liganded MurA has a more compact structure. The enzyme is made up of two globular domains, which are connected by a double-stranded hinge. Each domain is roughly spherical with a radius of about 20 Å. The main chain of each domain is very similar, with three parallel internal helices surrounded by three helices and three four-stranded β-sheets exposed to solvent. There is an approximate threefold symmetry, which relates secondary structure elements within each of the domains. Multiple forms of Gram-negative MurA have been crystallized to date. It was shown that UDP-NAG binding is responsible for the conformational change that is required for catalysis to proceed (46). Catalysis is blocked by exclusion of PEP from the active site by the covalent bond formation between fosfomycin and residue C115 of MurA (50). Mutation of this residue to the aspartic acid residue is the basis for fosfomycin resistance of *M. tuberculosis* (53).

METHODS

Overexpression and purification of Qβ A_2

A_2 forms inclusion bodies when overexpressed at 37º C. Previous attempts at obtaining purified A_2 were concentrated at finding a right detergent system. However, recently the soluble and active A_2 was purified as a fusion construct MBP-A_2. The 93 kDa fusion protein consists of an N-terminal his-tag, followed by MBP, a short TEV cleavage site-containing linker, and A_2. Expressed from a plasmid, A_2 will lyse *E. coli* cells in about 40 minutes. In order to protect the cells and obtain sufficient yield of the protein, MBP-A_2 was co-expressed along with *B. subtilis* MurA. Cold expression at 16º C resulted in mostly soluble protein. The complex was purified with amylose resin (NEB) yielding 10 mg of protein per liter of media grown. Functionality was determined by inhibition of MurA in a phosphate release assay; reaction conditions were previously described (54). The inhibition was somewhat lower than that observed by Bernhardt and colleagues (45) for an excess of purified phage particles inhibiting MurA in a crude lysate. One possibility could be attributed to the steric hindrance of MurA imposed by the bulky MPB fusion, as seen elsewhere (55).

From previous truncation analyses, several lytic fragments of A_2 were identified (4). The construct of a functional truncation MBP-$A_2$190 (residues 1-190) has been purified using the same procedure as for the full-length protein. The yield after purification was 0.5 mg per liter culture, significantly lower than that for the full-length protein.

To determine the overall structural fold of A_2, circular dichroism (CD) was performed on equimolar concentrations (2 μM) of MBP-A_2, MBP-$A_2$190, and MBP (Catrina Reed, unpublished). The CD spectrum of MBP has been published (56), as was its crystal structure (57). The results for the percent of alpha-helical content of the MBP control (36%) were within error of the published structure (40%). A_2 protein's alpha-helical content was determined to be about 26% by deconvolution analyses, whereas the $A_2$190 truncation was about 36% alpha-helical. Structure analysis (Jpred3) (58) predicted that both the full length A_2 and $A_2$190 were about one-third alpha-helical.

E. coli MurA-A_2 complex

Catrina Reed (Young lab) provided *in vitro* evidence for a direct interaction of A_2 and MurA with "pull aside" experiments. Amylose magnetic beads were used to sequester MBP-A2 to the side of the tube with application of a magnet. MurA associated with A_2 was eluted in the bound fraction. The data obtained suggest that MBP-A_2 preferentially binds to both liganded states of MurA, thus a closed conformational state of MurA is required for A_2 association (Figure 3.4).

For TEV cleavage, MBP-A_2, TEV, and *E. coli* MurA with UDP-NAG bound were incubated overnight at 4 ºC with no shaking. The mixture was then analyzed for soluble and insoluble fractions by Western blotting using antibodies directed against either MurA or A_2. In the absence of MurA, A_2 cleaved from MBP-A_2 fusion was insoluble; but in the presence of MurA with UDP-NAG, A_2 stayed soluble indicating the formation of an A_2-MurA complex.

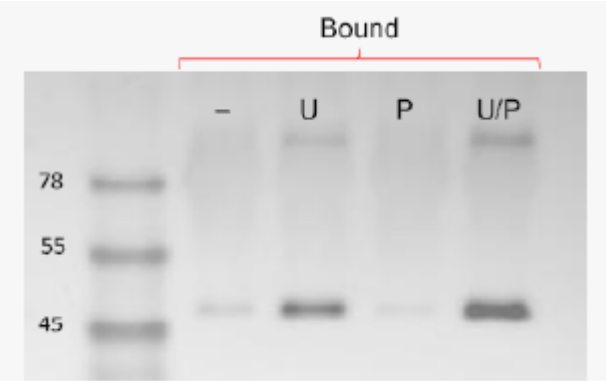


Figure 3.4 Western blot of MBP-A_2/MurA complex. Bound fraction were Western blotted with the MurA antibody. MurA bands are visible in UDP-NAG (U) and UDP-NAG/PEP (U/P) substrate bound fractions (courtesy of C. Reed).

Similar experimental setup using *Bacillus subtilis* MurA instead of *E. coli* MurA failed to show that A_2 is capable of binding to any form of *Bacillus subtilis* MurA.

Gel exclusion chromatography of the cleavage reaction will be performed to purify the complex followed by crystallization experiments. X-ray crystallography will be used in the investigation of the MurA—A_2 binding interface.

A fusion construct of the MurA-A_2 is being made by C. Reed. It consists of the N-terminal his-tag—MurA—a short linker with the TEV site—A_2. The idea is to have MurA close to the N-terminal (lytic) side of A_2. The construct will be assessed for expression levels and solubility.

The search for soluble lytic truncations of A_2 is in progress. Secondary structure analysis of A_2 predicted two alpha-helical stretches from residues 147-176 and 180-188. Functionality of the constructs will be assayed by culture lysis prior to purification. Overall protein fold and stability will be verified by SDS-PAGE and CD analyses.

Structural studies of a MurA complex with full length A_2 or a lytic fragment would ultimately provide detailed information about residues involved in binding. From these analyses generation of a lytic peptide might lead to the discovery of a novel "peptide antibiotic".

Purification, crystallization and data collection of *Bacillus subtilis* MurA

B. subtilis MurA was over expressed and purified in both the native form and as the selenomethionine derivative. pZE12 vector was used for the native overexpression after comparing favorably to different pET vectors both in terms of total cell mass accumulation and the expression levels. *B. subtilis* MurA was purified on the Talon immobilized metal affinity chromatography (IMAC) column followed by Superdex 75 size exclusion column. The purity of the protein was above 95% as judged by SDS-PAGE. MurA was concentrated to 10-12 mg/ml and crystallized using the hanging drop vapor diffusion method. Crystals appeared in about 3 days in several PEG 3350 based buffer conditions.

The native crystal diffracted up to 2.5 Å resolution at the APS 23ID source (Argonne Labs, IL). Despite its relatively high homology to either *E. coli* MurA or *Enterobacter clocae* MurA (around 49%), solving the structure of *B. subtilis* MurA by molecular replacement (MR) method proved difficult. Attempts were made using different models (whole molecules as well as the sub-domains) and different software programs including Phaser and MolRep.

In order to utilize the single anomalous dispersion (SAD) method, the selenomethionine analog of *B. subtilis* MurA was overexpressed using B834 (DE3) (methionine auxotroph) cell line from *Novagen* in pET11a vector under the T7 promoter. The protein was purified as described above. The crystallization conditions were re-screened against the entire crystallization library, as the original conditions did not produce high quality crystals. Protein was crystallized by the hanging drop method in HEPES buffered solution with 20% PEG 2000 and LDAO detergent. The data set was collected at the APS source 19ID (Argonne Labs, IL). The crystals diffracted to 2.3 Å resolution with unit cell parameters of a=b= 82.16 Å, c= 262.78 Å in tetragonal crystal system. Crystallographic data for Se-Met crystal is summarized in Table 3.1.

Table 3.1 Crystal Data and Refinement Statistics for *B. subtilis* MurA

Data Collection Statistics	
wavelength (Å)	0.97929
data set	peak
resolution (Å)	48.43-2.28
completeness % (last shell)	98.9 (88.6)
$I/\sigma(I)$ (last shell)	
no. of unique reflections	39799
R_{sym} (%)	
Refinement Statistics	
unit cell dimensions	a=b=82.161 Å, c=262.783 Å
space group	P 43 21 2
no. of protomers per asymmetric unit (Z)	2
no. of reflections in the working set	
completeness (%)	94.19
R_{factor} (%) / R_{free} (%)	23.6 / 27.6
no. of protein residues	834
no. of solvent molecules	135
average B factor for protein (Å)	60
rmsd for bond length (Å)	0.009
rmsd for bond angles (deg)	1.63

RESULTS AND DISCUSSION

The UDP-*N*-acetylglucosamine enolpyruvyl transferase (MurA) is an essential enzyme in the biosynthesis of the peptidoglycan layer of the bacterial cell wall. MurA catalyzes the transfer of enolpyruvate from phosphoenolpyruvate to UDP-*N*-acetylglucosamine (UNAG) to form enolpyruvyl-UDP-*N*-acetylglucosamine (EP-NAG), the first committed step in the bacterial cell wall synthesis. Previously published crystal structures of MurA revealed that the active site loop can adopt a closed conformation upon binding of its substrate UNAG (based on *E. coli*:UNAG and *E. coli*:fosfomycyn binary complex structures) or an open conformation in the ligand free state (based on apo *E. Cloacae* structure). A_2, the maturation protein of bacteriophage Qβ inhibits *E.coli* MurA. Previous studies have suggested that A_2 should bind to the closed form of MurA and block the substrate access route, possibly by locking the active site loop in the closed conformation. Perhaps, A_2 can preferably bind only to the substrate bound closed form of MurA. It has also been recently shown that A_2 does not associate with *B. subtilis* MurA *in vitro* even in presence of ligands (5). Since the structures of *E. coli* MurA in the UNAG bound state is available, the structure of *B. subtilis* MurA (Bs-MurA here after), the first structure from a Gram-positive organism, preferably in the closed form, allows for the direct comparison between these two homologs.

To provide further insight in to the molecular basis of A_2-MurA interactions, we have determined the X-ray crystal structure of Bs-MurA. The structure of Bs-MurA with

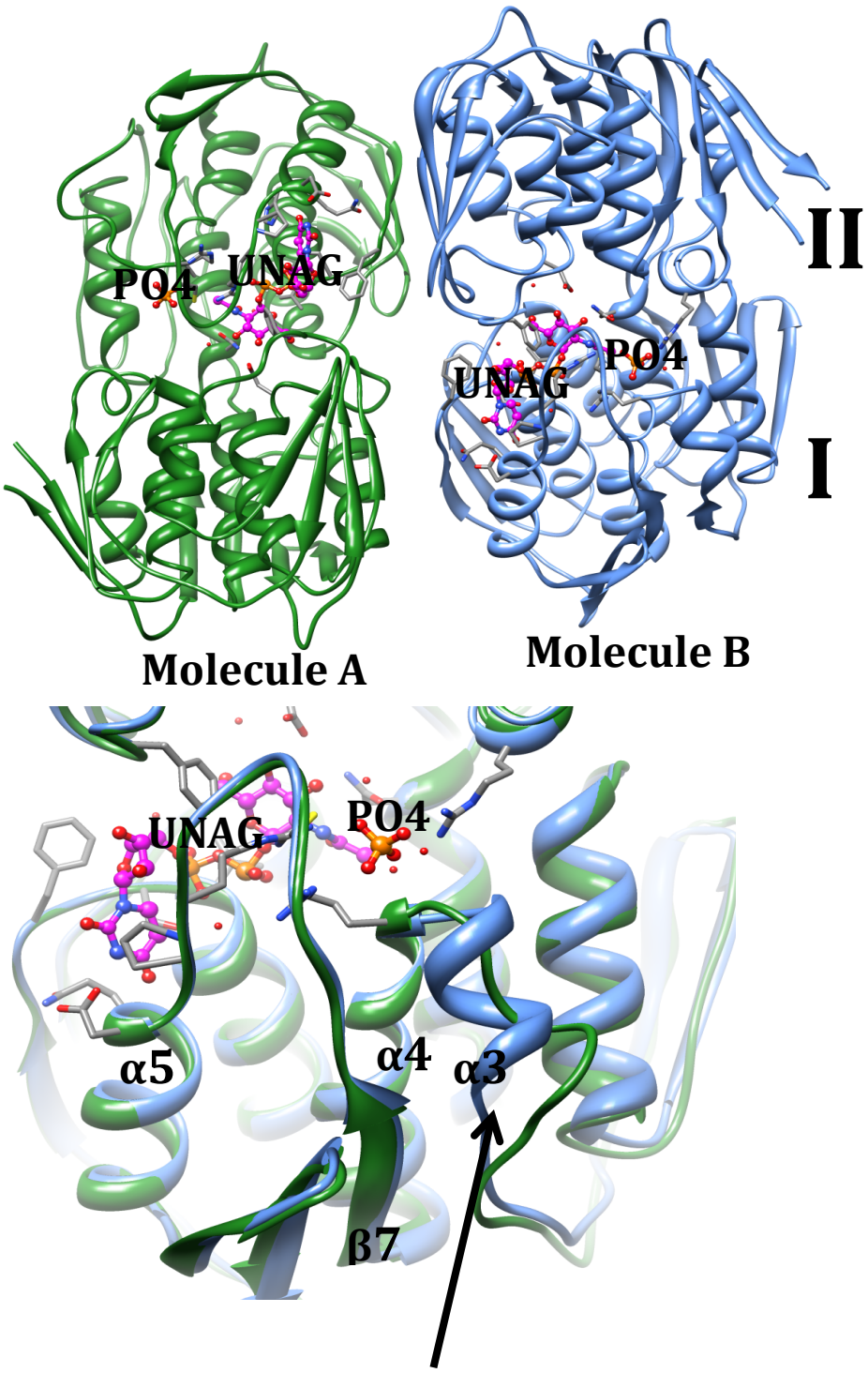


Figure 3.5 ASU of MurA. Two molecules Bs-MurA in ASU. Region of structural difference indicated.

two molecules in the asymmetric unit (molecule A and B) was solved using the SAD method with crystals of selenomethionylated protein formed in the space group $P4_32_12$ at 2.3-Å resolution. The SAD electron density map was of moderate quality and allowed the 60 % of the polypeptide to be traced for both the molecules. Compared to molecule B of the asymmetric unit, molecule A contained more disordered electron density regions. Several cycles of manual model building guided through the *E. coli* model, SAD experimental map and shake 'n' warp averaged map yielded the final model of Bs-MurA with two molecules in the asymmetric unit as shown in Figure 3.5. Interestingly, continuous additional electron density was clearly visible in the active sites of both molecules of the asymmetric unit in all stages of refinement, at 1σ cutoff (Figure 3.6). This extra electron density, from an unbiased electron density map, was very clear and could be readily fitted and refined by the substrate UNAG (Figure 3.6). Significant electron density was also visible near the Cys117 residue and could be modeled and refined by a phosphate ion, similar to previously observed MurA binary complex structures. However, we never included either the UNAG substrate or the phosphate ion in our crystallization conditions (see methods section). It is most likely that the enzyme captured these ligands during the process of *E coli* over expression. Further refinement after including these molecules yielded the final model with an R_{factor} of 23.7% and an R_{free} of 26.7 % (Table 1). The final model had good stereochemistry for both molecules as determined by the Ramachandaran plot. The structure reveals that there is no notable difference between two molecules of the asymmetric unit (Figure 3.5). This is illustrated in Figure 3.5 (close up), showing the superimposition of the individual molecules.

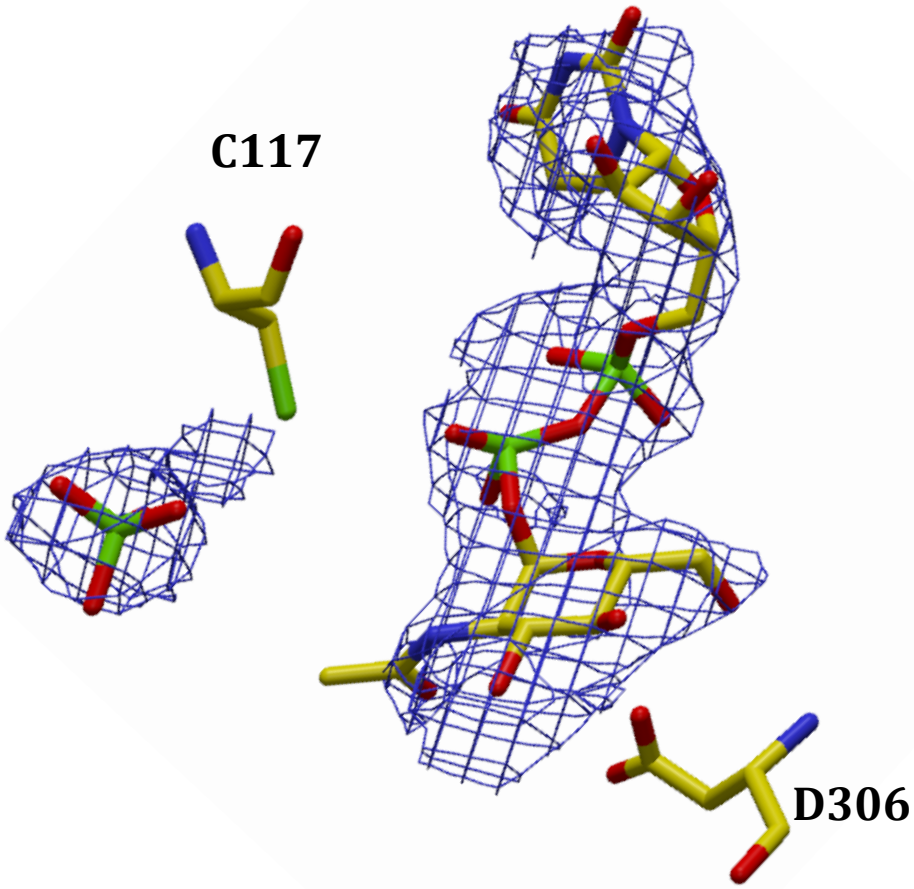


Figure 3.6 Electron density of substrate. Shake'n'warp electron density of UNAG. The substrate and phosphate bound in the catalytic cleft shown.

Molecule A of the asymmetric unit of Bs-MurA superimposes on molecule B with an r.m.s.d. of 0.52 Å. Molecule A showed some differences near the α3 helix compare to molecule B, However, the rest of the structure of molecule A including the active site loop is very similar to molecule B. We have selected the more ordered molecule, molecule B of the Bs-MurA structure for further analysis.

Overall structure of Bs-MurA

Like the *E. coli* MurA crystal structure, each subunit of Bs-MurA contains two similarly folded globular domains called domain I (residues 21 to 228) and domain II (residues 1 to 20 and 230 to 417). The two domains are connected through a double linker (β1-α1 and β14-α8) at the inner most part of the active sited. The internal arrangement of both the domains is similar with the previously defined "two-α-helix, four- β-strand" motif repeated three times around an approximate 3-fold symmetry axis. In each domain, three beta sheets form a triangular periphery to capture a bundle of 6 helices inside. In addition the structure also consists of two small α helices. The complete nomenclature is shown in Figure 3.7.

The active site of Bs-MurA

Each molecule of the *Bs*-MurA:UNAG binary complex structure contained one UNAG molecule, similar to previously observed conformation (coordinates taken from PDB ID: 3KQJ). The UNAG binding active site of *Bs*-MurA is located in a deep pocket

between the two domains, enclosed by the β7-α5 connecting loop (referred as active site loop in previous publications). Amino acids from both domains contribute to

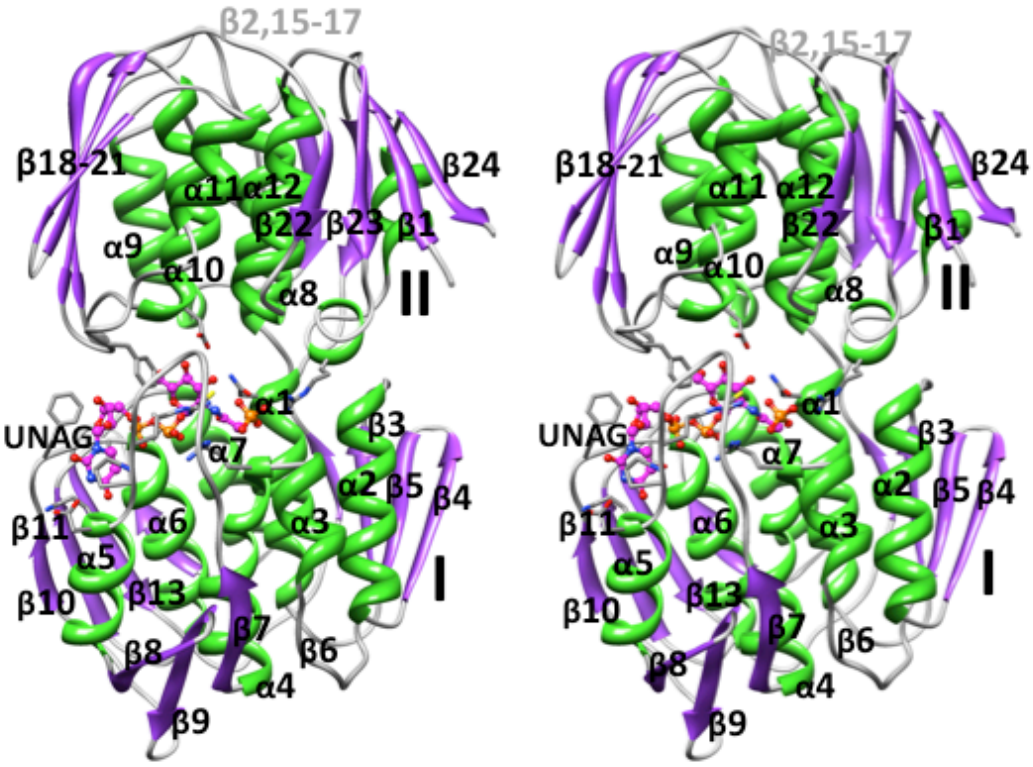
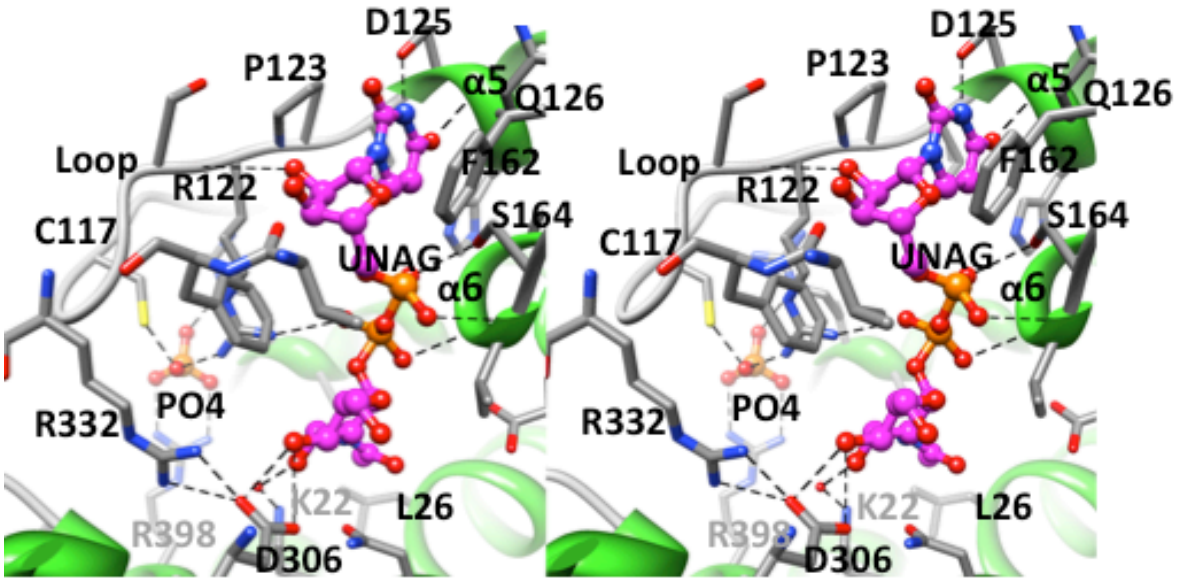


Figure 3.7 Bs-MurA nomenclature (stereo view).



ASP 125.B OD1 <-> UD1 950.A N3	2.696
GLN 126.B N <-> UD1 950.A O4	2.904
SER 164.B OG <->UD1 950.A O2A	2.791
VAL 165.B N <-> UD1 950.A O1A	2.733
GLY 166.B N <-> UD1 950.A O1B	3.082
UD1 950.A O4' <-> ASP 306.B OD2	3.006
UD1 950.A O4' <-> ASP 306.B OD1	2.809
ARG 122.B NH1 <-> UD1 950.A O2B	2.969
ARG 398.B NH2 <-> PO4 921.A O4	2.731
ARG 398.B NH1 <-> PO4 921.A O2	2.624
PO4 921.A O1 <-> ARG 122.B NH2	3.093
PO4 921.A O3 <-> ARG 93.B NE	2.996
CYS 117.B SG <-> PO4 921.A O1	2.788
ARG 332.B NH2 <-> ASP 306.B OD1	2.791
ARG 332.B NH1 <-> ASP 306.B OD1	2.659
UD1 950.A O2' <-> SER 121.B O	2.958

Figure 3.8(A) Binding pocket close up. Interactions of UNAG in the binding pocket (stereo view). Hydrogen bond distances (Å) for UNAG interactions with residues of MurA given.

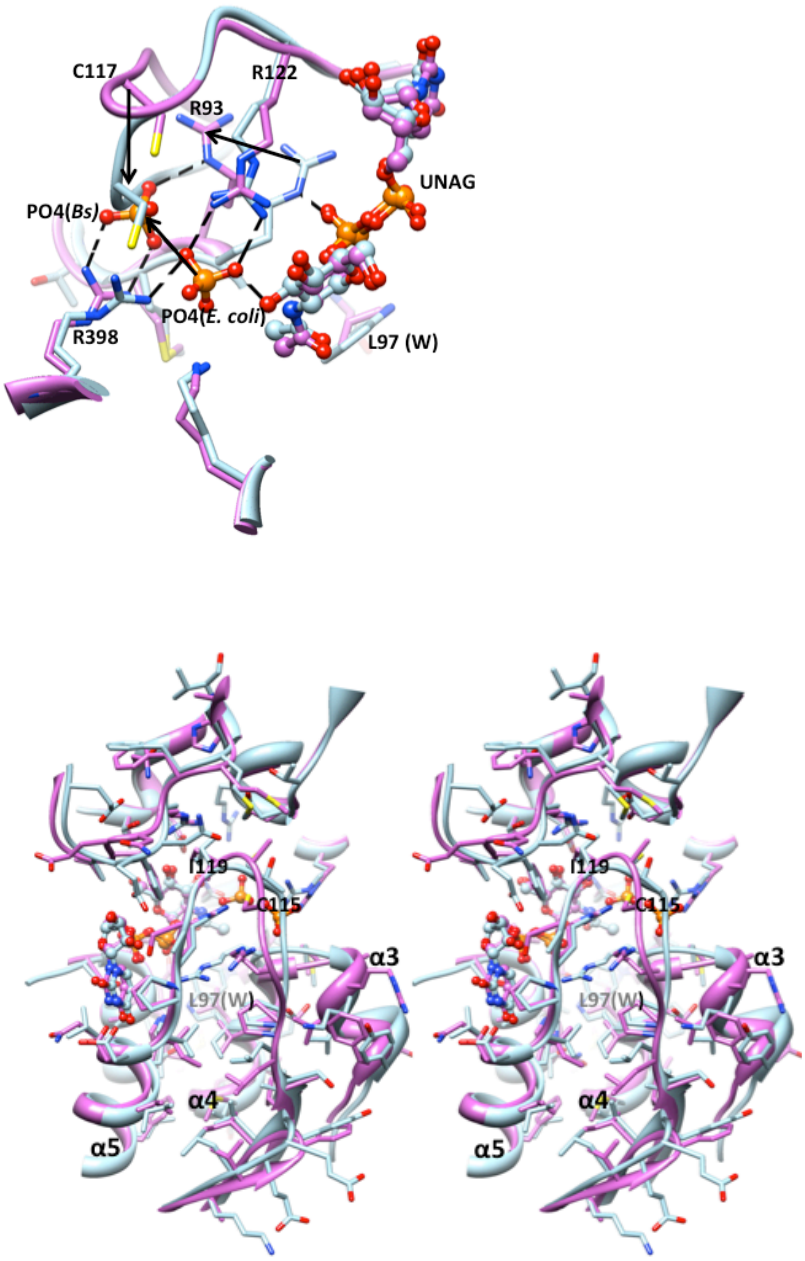


Figure 3.8(B) Binding pocket close up. UNAG in the binding pocket.

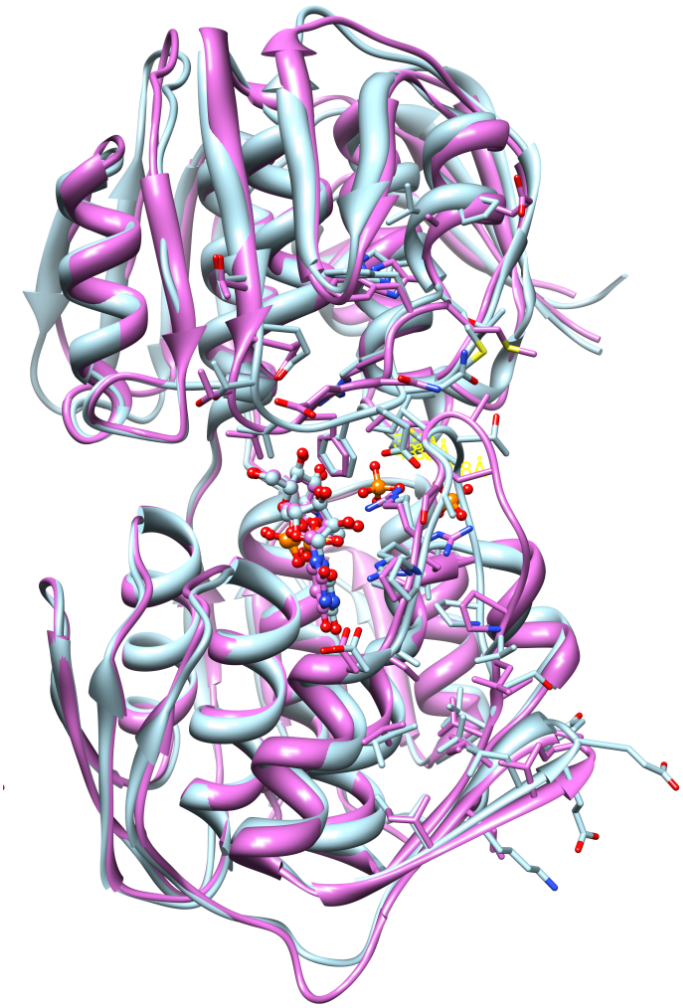


Figure 3.9 MurA superposition. Bs-MurA (purple) superimposed over *E. Coli* MurA (PDB: 3KQJ). Both structures have UNAG bound.

the binding of UNAG (Figures 3.8A,B and 3.9). The MurA-UNAG interactions for the uridine and ribose moieties of UNAG are very similar with *E. coli*:UNAG binary complex structure. The uridinyl ring of UNAG packed against the two hydrophobic patches formed by Pro123 on one side and Gln126 (Leu124 in the *E. coli* structure) on the other side. It also forms hydrogen bond with the side chain oxygen atom of Asp125 and main chain carbonyl oxygen atom of Gln126.

The ribose moiety of the UNAG forms one hydrogen bond with the main chain carbonyl oxygen atom of Val 328 (of β19-α11 connecting loop) and the pyrophosphate bridge is involved in hydrogen bonds with the main chain nitrogen atom of Val165 and Gly166, the side chain hydroxyl of Ser164, and the side chain NH1 atom of Arg122 (of the catalytic loop). The N-acetylglucosamine group (NAG) forms a water mediated hydrogen bond with Lys22 and two direct hydrogen bonds with Asp306. Distances of all these hydrogen bonds are listed in Figure 3.8A. In addition two hydrogen bonds, hydrophobic interactions also play a role in binding of UNAG to Bs-MurA, similar to *E. coli* MurA. Phe329 inserts its side chain close to the NAG moiety to interact with the C5' and C6' carbon atoms of NAG group. Phe329 also interacts with the C3 carbon atom of the ribose group. The terminal C8' carbon atom of NAG forms hydrophobic interactions with the Leu26 residue of Bs-MurA, similar to *E. coli* MurA. However, the Trp residue of *E. coli* is mutated into Leu97 in Bs-MurA and does not interact with UNAG. This Trp to Leu change may be the initiation point for the loop movement observed in the Bs-MurA compared to *E. coli* MurA (see next paragraph, Figure 3.10, and next section).

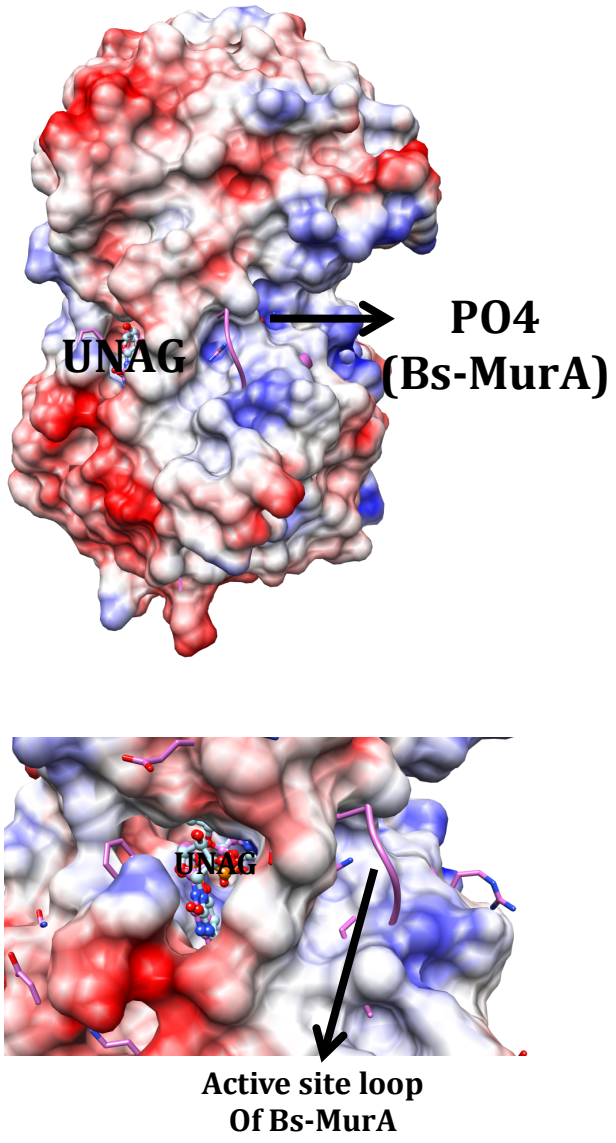


Figure 3.10 Loop arrangements. Bs-MurA loop protrudes beyond *E. Coli* MurA surface.

The NH2 atom of Arg122 hydrogen bonds or salt bridges (depending on the protonation state of the phosphate ion) to the nearby oxygen atom of a free phosphate ion observed near the Cys117 of the active site loop. Like UNAG binding, the presence of phosphate ion near the active site is also unexpected, as we never included the ligands in the crystallization condition. It is likely that during the overexpression in the *E. coli* cells the Bs-MurA captured the UNAG and the phosphate ion. We also checked the possibility of the product EP-UNAG in the place of UNAG and over ruled it based on the fit to the electron density map. The free phosphate ion is located 10 Å from the UNAG's pyro phosphate group and make two hydrogen bonds with the gunadinium nitrogen atoms of Arg398 and NE atom of the Arg93 atom. In *E. coli*, this phosphate hydrogen bonds with the tail end of the UNAG substrate, in contrast, it is moved 5 Å more towards the active site loop. The active site loop of Bs-MurA, also adopts different conformation (Figure 3.8B). Interestingly, the side chain of Arg93 needs to adopt a different conformation compare to that of *E. coli* MurA to interact with the phosphate oxygen atom. More importantly, the phosphate oxygen atom also coordinates with side chain of the highly conserved Cys117. The position of the active site loop, Cys117 and the free phosphate ion all are strikingly different compared to that of *E. coli* binary complex structure (Figure 3.8B), while the UNAG position and interacting residues are mostly similar.

The active site loop of Bs-MurA adopts different conformation compare to that of *E. coli* MurA. This conformational change is augmented by several mutations in the

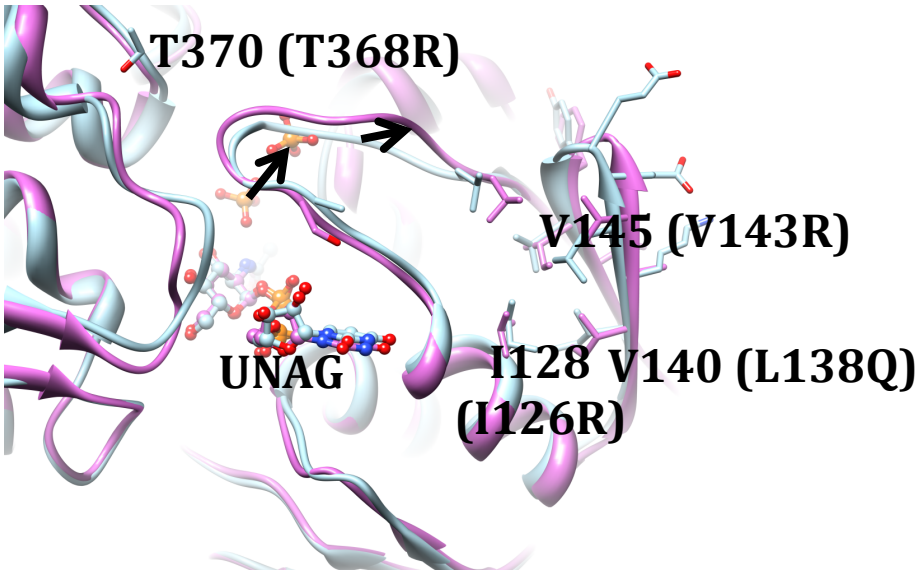


Figure 3.11 Mapping of the RAT mutants. Bs-MurA catalytic loop position relative to *E. coli* MurA. RAT mutants shown with corresponding Bs-MurA residues.

sequence of Bs-MurA compare to *E. coli* MurA, starting from Leu97 (*Trp in E. coli*) (see Figure 3.8B). The overall protein fold of the Bs-MurA is very similar to that of *E.coli*:UNAG complex (with an rms difference of 0.87 Å (for 380 residues) after the superposition of the main chain Cα atoms), except the region around the Leu97 residue, adjacent α3 helix and the active site loop near α3 helix (Figure 3.10).

Therefore, it is evident that Bs-MurA:UNAG complex structure adopts a closed conformation similar to previous ligand bound MurA structures. However, compared to *E. coli* MurA:UNAG structure, the active site loop of Bs-MurA adopts a new conformation augmented by the sequence variation of this Gram-positive protein. A_2, the maturation protein of bacteriophage Qβ inhibits *E.coli* MurA. Previous studies have suggested that A_2 should bind to the closed form of MurA and block the substrate access route. However, we have no idea about the structure of A_2 or its anchoring points in MurA. We compared the ligand bound closed form of Bs-MurA (not recognized by A_2) crystal structure with the *E. coli* MurA:UNAG closed form structure (the preferred partner of A_2). As shown in Figures 3.9, 3.10,and 3.11 the main differences are near the active site loop region. The catalytic loop of Bs-MurA protrudes out more compared to that of *E. coli* structure, possibly disrupting the anchoring points for the tight binding of A_2. It is possible that the extra protrusion blocks the binding of A_2 to Bs-MurA. This notion is consistent with the fact that mutations in *E. coli* MurA in this region compromise the ability of A_2 to inhibit *E. coli* MurA (Figure 3.11).

Electrostatic surface comparison of Bs-MurA and *E. coli* MurA

Comparison of electrostatic potential surfaces between Bs-MurA and *E. coli* MurA revealed the following key information (Figure 3.12). The substrate-binding site of *E. coli* MurA is significantly larger than that of Bs-MurA and it exhibits the higher positive charge distribution. This is in agreement with the activity assay data, suggesting higher enzymatic activity of *E. coli* MurA.

In context of the preferential binding of A_2 maturation protein to *E. coli* MurA, the electrostatic surface comparison of the two enzymes possibly offers an additional clue. In addition to the active site loop conformation, discussed earlier, the local change in the charge distribution near the active site loop and more dominant RAT mutants of *E. coli* MurA (indicated by arrows in Figure 3.12) could potentially play a role in the difference of A_2 binding.

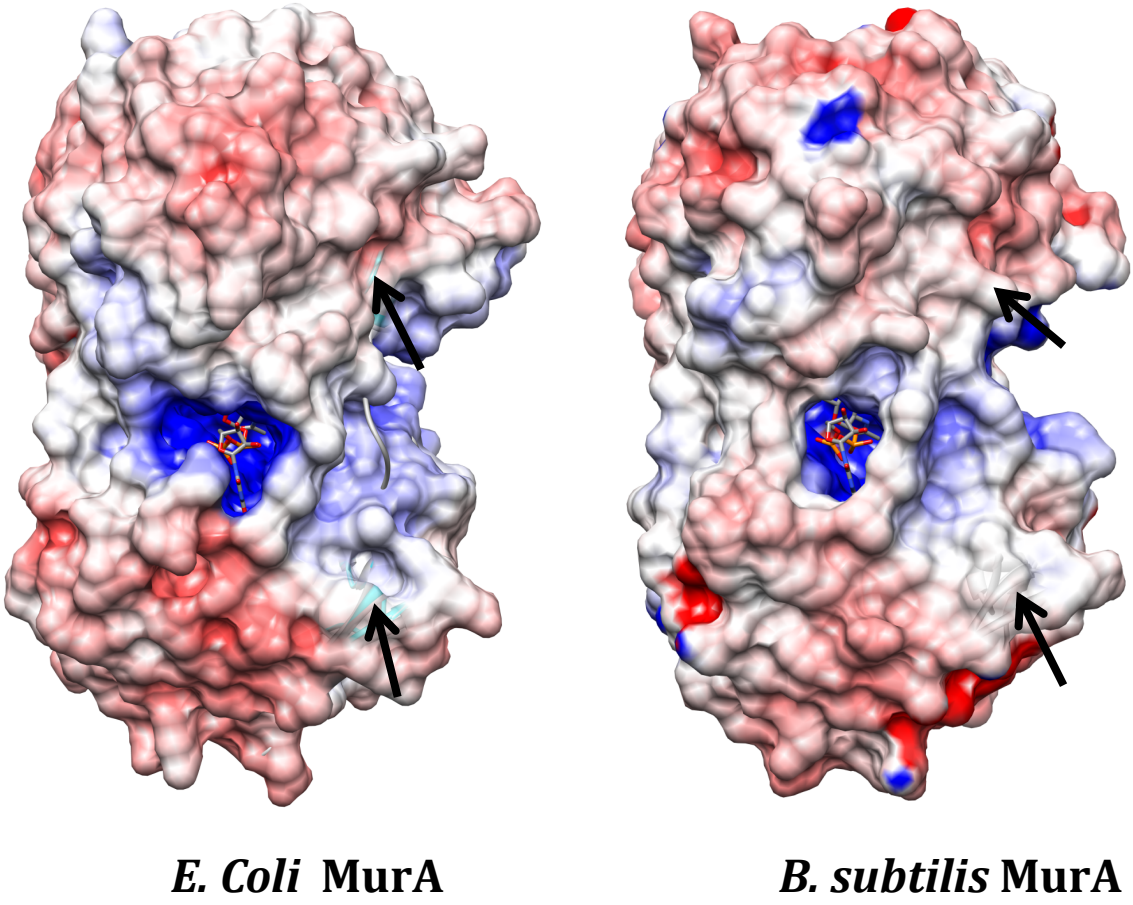


Figure 3.12 Electrostatic surface potential of *E. coli* MurA and *B. subtilis* MurA molecules (APBS server).

CHAPTER IV

CONCLUSIONS AND FUTURE DIRECTIONS

CHAPTER II SUMMARY AND FUTURE DIRECTIONS

Periplasmic domains of antiholin RI, and holin T of phage T4 were overexpressed. Two atomic-resolution crystal structures were presented: a homo oligomeric assembly of RI and an RI-T complex that crystallized as a hetero tetramer. Based on the crystal structures, a number of important observations concerning the process of lysis inhibition (LIN) and of holin-dependent lysis in general could be made. Antiholin RI could exist as a dimer, or even a tetramer, in its extended conformation, and as a monomer in its compact form. Helix H1 of RI is domain swapped between the monomeric and oligomeric conformations. Only RI monomer is capable of binding the holin and, based on the crystal contacts in holin-antiholin complex, RI must be in its compact monomeric conformation to bind T. Periplasmic domain of holin T is a GAF fold and might be capable of binding a small molecule or nucleotide. RI-T complex is a hetero tetramer and each RI contacts both T molecules. Binding of RI separates α1, α3 helices of T molecules in the complex by 28 Å. In absence of RI, holin T can dimerize through its (α1, α3) x2 helical bundle.

The crystal structure of antiholin RI shows a homo oligomeric assembly. This finding is collaborated by both the computational analysis program PISA and by the structural homology search with DALI and VAST databases. Additional experiments

can be designed to further test the oligomeric state of RI. Small-angle scattering of X-rays (SAXS) is an established method for the low-resolution structural characterization of biological macromolecules in solution and should be applicable in this case. The technique allows one to assess the oligomeric state of the proteins and protein complexes. Additional experiments such as analytical ultracentrifugation (AUC) and cross-linking can also be designed.

As previously mentioned, the periplasmic domain of holin T alone was not crystallized so far due to its propensity to oligomerize and precipitate out of solution at concentrations usually required for protein crystallization. Since T is a GAF domain and there are a number of structures of GAF domains in the PDB, one can design a mutant of T to be crystallized as a dimer.

The majority of GAF folds are expected to bind small molecules or nucleotides. Analysis of the crystal structure of T confirms that T is at least capable of binding a small molecule based on loop arrangements and key residues at the catalytic site. A number of experiments can be designed to test this. One way is to do a metabolite screening to search for a substrate. T could also be designed with a cGMP-binding motif at the expected catalytic site. Binding of cGMP would make T resistant to antiholin RI. It is known that the periplasm is deficient in small molecules because of the 600 Da general OMP porins. If the GAF fold of T does bind a small molecule, it might be dNMPs produced from likely degradation of T4 DNA during LIN.

CHAPTER III SUMMARY AND FUTURE DIRECTIONS

Crystal structure of *B. subtilis* MurA was solved at an atomic resolution. This is the first such structure for a MurA from a Gram-positive organism. Since the crystal structures of MurA from Gram-negative organisms are already known, this allows for the direct comparison of Gram-positive and Gram-negative homologs. The structure of *B. subtilis* MurA was compared to that of *E. coli* MurA and major structural differences were identified.

The structure of *B. subtilis* MurA was evaluated with respect to its ability to resist inhibition by A_2, the maturation protein of phage Qβ. It appears that the more flexible catalytic loop of *B. subtilis* MurA protrudes farther out compared to the loop of *E. coli* MurA and creates enough hindrance to prevent A_2 from establishing secure contact points. Also, there is a slight difference in the electronegative surface at the expected A_2 binding site that could potentially contribute to resistance.

Crystallization of A_2, possibly as a complex with *E. coli* MurA, remains a challenge.

The original RAT mutant of *E. coli* MurA, L138Q, could be crystallized and is expected to have its catalytic loop protruding similarly to one of *B. subtilis* MurA.

V. cholerae MurA, which is very close to *E. coli* MurA on the phylogenetic tree, could also be crystallized. It is expected to be inhibited by A_2. *V. cholerae* MurA is an attractive drug target in light of recent major outbreaks of drug-resistant cholera in the world.

REFERENCES

1. Bernhardt, T. G., Wang, I. N., Struck, D. K., and Young, R. (2002) *Res Microbiol* **153**, 493-501

2. Benzer, S. (1961) *Proc Natl Acad Sci U S A* **47**, 403-415

3. Savva, C. G., Dewey, J. S., Deaton, J., White, R. L., Struck, D. K., Holzenburg, A., and Young, R. (2008) *Mol Microbiol* **69**, 784-793

4. Langlais, C. (2007) Doctoral dissertation. Department of Biochemistry and Biophysics, Texas A&M University, College Station, TX

5. Reed, C. (2009) Personal communication. Graduate research assistant, Department of Biochemistry and Biophysics, Texas A&M University, College Station, TX

6. Leiman, P. G., Kanamaru, S., Mesyanzhinov, V. V., Arisaka, F., and Rossmann, M. G. (2003) *Cell Mol Life Sci* **60**, 2356-2370

7. Goldberg E., G. L., and Letellier L. (1994) *Molecular Biology of Bacteriophage T4*, ASM Press, Washington DC

8. Baschong, W., Aebi, U., Baschong-Prescianotto, C., Dubochet, J., Landmann, L., Kellenberger, E., and Wurtz, M. (1988) *J Ultrastruct Mol Struct Res* **99**, 189-202

9. Crowther, R. A., Lenk, E. V., Kikuchi, Y., and King, J. (1977) *J Mol Biol* **116**, 489-523

10. Kanamaru, S., Leiman, P. G., Kostyuchenko, V. A., Chipman, P. R., Mesyanzhinov, V. V., Arisaka, F., and Rossmann, M. G. (2002) *Nature* **415**, 553-557

11. Kanamaru, S., Gassner, N. C., Ye, N., Takeda, S., and Arisaka, F. (1999) *J Bacteriol* **181**, 2739-2744

12. Simon, L. D., and Anderson, T. F. (1967) *Virology* **32**, 298-305

13. Miller, E. S., Kutter, E., Mosig, G., Arisaka, F., Kunisawa, T., and Ruger, W. (2003) *Microbiol Mol Biol Rev* **67**, 86-156

14. Cerritelli, M. E., Wall, J. S., Simon, M. N., Conway, J. F., and Steven, A. C. (1996) *J Mol Biol* **260**, 767-780

15. Bhattacharyya, S. P., and Rao, V. B. (1993) *Virology* **196**, 34-44

16. Leffers, G., and Rao, V. B. (2000) *J Biol Chem* **275**, 37127-37136

17. Tran, T. A., Struck, D. K., and Young, R. (2005) *J Bacteriol* **187**, 6631-6640

18. Weaver, L. H., and Matthews, B. W. (1987) *J Mol Biol* **193**, 189-199

19. Berry, J., Summer, E. J., Struck, D. K., and Young, R. (2008) *Molecular Microbiology* **70**, 341-351

20. Dressman, H. K., and Drake, J. W. (1999) *J Bacteriol* **181**, 4391-4396

21. Hershey, A. D. (1946) *Genetics* **31**, 620-640

22. Cairns, J., G.S. Stent, J.D. Watson. (1966) *Phage and the Origins of Molecular Biology*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

23. Benzer, S. (1955) *Proc Natl Acad Sci U S A* **41**, 344-354

24. Alikhanian, S. I., Grinberg, K. N., Krylov, V. N., Maisurian, A. N., and Oganesian, M. G. (1965) *Izv Akad Nauk SSSR Biol* **4**, 542-549

25. Tran, T. A., Struck, D. K., and Young, R. (2007) *J Bacteriol* **189**, 7618-7625

26. Dewey, J. S., Savva, C. G., White, R. L., Vitha, S., Holzenburg, A., and Young, R. (2010) *Proc Natl Acad Sci U S A* **107**, 2219-2223

27. White, R., Chiba, S., Pang, T., Dewey, J. S., Savva, C. G., Holzenburg, A., Pogliano, K., and Young, R. (2011) *Proc Natl Acad Sci U S A* **108**, 798-803

28. Paddison, P., Abedon, S. T., Dressman, H. K., Gailbreath, K., Tracy, J., Mosser, E., Neitzel, J., Guttman, B., and Kutter, E. (1998) *Genetics* **148**, 1539-1550

29. Krissinel, E., and Henrick, K. (2007) *J Mol Biol* **372**, 774-797

30. Krissinel, E. (2010) *J Comput Chem* **31**, 133-143

31. Krissinel, E. (2007) *Bioinformatics* **23**, 717-723

32. Krissinel, E. (2011) *Acta Crystallogr D Biol Crystallogr* **67**, 376-385

33. McGrath, M. E., Gillmor, S. A., and Fletterick, R. J. (1995) *Protein Sci* **4**, 141-148

34. McGrath, M. E., Hines, W. M., Sakanari, J. A., Fletterick, R. J., and Craik, C. S. (1991) *J Biol Chem* **266**, 6620-6625

35. Yang, S. Q., Wang, C. I., Gillmor, S. A., Fletterick, R. J., and Craik, C. S. (1998) *J Mol Biol* **279**, 945-957

36. Bennett, M. J., Schlunegger, M. P., and Eisenberg, D. (1995) *Protein Sci* **4**, 2455-2468

37. Eggers, C. T., Wang, S. X., Fletterick, R. J., and Craik, C. S. (2001) *J Mol Biol* **308**, 975-991

38. Gillmor, S. A., Takeuchi, T., Yang, S. Q., Craik, C. S., and Fletterick, R. J. (2000) *J Mol Biol* **299**, 993-1003

39. Tarricone, C., Xiao, B., Justin, N., Walker, P. A., Rittinger, K., Gamblin, S. J., and Smerdon, S. J. (2001) *Nature* **411**, 215-219

40. Kwon, E., Kim, D. Y., Gross, C. A., Gross, J. D., and Kim, K. K. (2010) *Protein Sci* **19**, 2252-2259

41. Sun, Q., Kuty, G. F., Arockiasamy, A., Xu, M., Young, R., and Sacchettini, J. C. (2009) *Nat Struct Mol Biol* **16**, 1192-1194

42. Young, I., Wang, I., and Roof, W. D. (2000) *Trends Microbiol* **8**, 120-128

43. Winter, R. B., and Gold, L. (1983) *Cell* **33**, 877-885

44. Weiner, A. M., and Weber, K. (1971) *Nat New Biol* **234**, 206-209

45. Bernhardt, T. G., Wang, I. N., Struck, D. K., and Young, R. (2001) *Science* **292**, 2326-2329

46. Schonbrunn, E., Svergun, D. I., Amrhein, N., and Koch, M. H. (1998) *Eur J Biochem* **253**, 406-412

47. Schonbrunn, E., Eschenburg, S., Krekel, F., Luger, K., and Amrhein, N. (2000) *Biochemistry* **39**, 2164-2173

48. Bugg, T. D., and Walsh, C. T. (1992) *Nat Prod Rep* **9**, 199-215

49. Schonbrunn, E., Sack, S., Eschenburg, S., Perrakis, A., Krekel, F., Amrhein, N., and Mandelkow, E. (1996) *Structure* **4**, 1065-1075

50. Skarzynski, T., Mistry, A., Wonacott, A., Hutchinson, S. E., Kelly, V. A., and Duncan, K. (1996) *Structure* **4**, 1465-1474

51. Eschenburg, S., Kabsch, W., Healy, M. L., and Schonbrunn, E. (2003) *J Biol Chem* **278**, 49215-49222

52. Skarzynski, T., Kim, D. H., Lees, W. J., Walsh, C. T., and Duncan, K. (1998) *Biochemistry* **37**, 2572-2577

53. Kim, D. H., Lees, W. J., Kempsell, K. E., Lane, W. S., Duncan, K., and Walsh, C. T. (1996) *Biochemistry* **35**, 4923-4928

54. Marquardt, J. L., Siegele, D. A., Kolter, R., and Walsh, C. T. (1992) *J Bacteriol* **174**, 5748-5752

55. Sorensen, H. P., and Mortensen, K. K. (2005) *Microb Cell Fact* **4**, 1-8

56. Ganesh, C., Shah, A. N., Swaminathan, C. P., Surolia, A., and Varadarajan, R. (1997) *Biochemistry* **36**, 5020-5028

57. Spurlino, J. C., Lu, G. Y., and Quiocho, F. A. (1991) *J Biol Chem* **266**, 5202-5219

58. Cole, C., Barber, J.D., Barton, G.J. (2008) *Nucleic Acids Res* **36**, 197-201

VITA

Name:	Vladimir Borisovich Kuznetsov
Address:	Texas A&M University, ILSB 301 Old Main, Room 2138 College Station, TX 77843
Email Address:	vbk@tamu.edu
Education:	B.S., Chemistry, University of California, Santa Cruz, 2000